



# Machine Learning Approaches to Predict Movie Success

Md. Afzazul Hoque and Md. Mohsin Khan<sup>(✉)</sup>

International Islamic University Chittagong, Chittagong, Bangladesh  
ete.mohsin@gmail.com

**Abstract.** The motion picture industry, commonly known as the film industry, is a huge investment sector. Furthermore, there is a massive amount of data related to movies on the internet. As a result, it became an interesting topic for research and analysis. Machine learning is a novel approach to data analysis. Our thesis presents a decision support system for the movie investing sector that employs various machine learning methods. In this context, our approach will assist film investment industries in avoiding investment risks. The algorithm evaluates historical data from numerous sources such as Rotten Tomatoes, IMDb, Box Office Mojo, and Meta Critics to calculate an estimated value of a film's success rate. The model will forecast movie success based on specific criteria such as the cast and directors, budget, month of release, runtime, several sorts of movie ratings, and movie reviews, and then process that data for categorization using machine-learning algorithms and a few other techniques.

**Keywords:** Film Industry · Machine Learning · Movie Success

## 1 Introduction

The motion picture industry, commonly known as the film industry, is a dominant market in the entertainment business. Each year more than 20 thousand movies are produced, distributed, and viewed worldwide in various languages, according to the Internet Movies Database (IMDB). These figures demonstrate the depth of the industry's impact on the market. In addition, to the cultural and socio-economic impact of the film, Motion Picture nowadays occupies a significant part of the business market, generating an average of 10 billion in annual growth, according to Boxofficemojo. The production houses invest much money in movie making every year. So, from this point of view, the film industry is a huge investment sector. However, larger business areas are more complex and difficult to choose how to invest. Big investments bring high risk. We cannot guess how a movie will do in the marketplace until the film opens in a darkened theatre. As film industry is getting competitive field and peoples are investing huge amount, so they need to are their film is going to gross or not. If we can predict them the result, it will help them a lot. As the film industry is expanding daily, a massive amount of data is now available on the web. Moreover, for this, it became an exciting area for data analysis. Predicting

a movie, whether it is successful or flopped, is not an easy task to do. The concept of “movie success” is quite different here; some films are considered successful based on their worldwide gross income, while others may not perform well in terms of box office revenue yet receive positive reviews and popularity. Many films did not make a lot of money when they were first released but became popular after a few years. For example, in 1999, David Fincher directed “Fight Club,” a widely popular film now. This film features actors such as Brad Pitt and Edward Norton. According to IMDb, however, this picture was a financial flop. The investment in “Fight Club” was 63 million USD. However, the global gross profit was only around 100 million USD, resulting in a net profit of only 37 million USD. That is not a very strong profit margin. However, “Fight Club” is a highly popular film right now, and every movie enthusiast is familiar with it. If profit is used as a criterion for success, then “Fight Club” is a financially flop movie. But if we look at other aspects, everyone can consider this film a successful film. As a result, it is a highly complicated issue for investors to make the right decision for these unforeseen circumstances of the film’s success. Research studies say that nearly 25% of film revenue comes during the first and second week of release [22]. As a result, predicting a movie’s success prior to its release is challenging.

In this research, we have attempted to construct a model that may enable investors to lower the risks in their investment. This research will be significant to the whole film business. A large portion of emerging artists can’t create films for investment since no investor is willing to invest in them. Investors have their reasons as well; not every investor has the confidence to invest in a movie with a newbie filmmaker since he/she does not have any prior work or enough experience to demonstrate. However, they are tremendously talented and enthusiastic about film creation. The forecast will aid an investor in choosing whether he/she wants to invest in new artists. It would be wonderful for a new artist in the market. The film industry is contributing significant money to the worldwide market. Suppose any new artists can produce movies without difficulties. In that case, it will inspire more artists, and as a consequence, more films will be created, and thus, the film industry’s contribution to the worldwide market will go up. Our purpose is to support young artists so that an investor can easily make investment decisions. However, our vision or significant goal is to do something for Bangladesh’s film industry in the future. There are a lot of enthusiastic young artists in Bangladesh who are amazingly passionate and dedicated to making new films but cannot do anything because of money; no production companies are willing to invest money blindly for them. We expect this research will benefit the Bangladeshi film industry because there are few production houses in our country, and producers lack the courage and financial resources to work with young artists. It will be extremely beneficial to them if we can offer them an idea of how a movie can do business after its release. We focused on only foreign movies when we created our dataset because we do not have enough data on Bangladeshi movies available on the web. Managing the film data of our country will be time-consuming. Furthermore, that is our future goal to help the film industry of Bangladesh by collecting data and making a prediction model.

For 838 movies, we utilized two features in our proposed system: pre-release features and post-release features. We can only forecast upcoming films by looking at pre-release features. Both pre-release and post-release features will perform forecasting immediately

after release. Five pre-release features and five post-release features are included. More features were included in this study to enable us to create a better and more standardized forecast. Instead of predicting flops and blockbusters solely, we will look at a variety of factors. [10]. However, we classify films into five categories depending on their box office revenue, from failure to blockbuster. Machine learning techniques such as Naive Bayes, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree, and Logistic Regression are available for multiclass prediction. These classifiers are good enough for binary classification. We applied all the classification algorithms to our dataset for prediction. Among those, Random Forest showed 94.76% accuracy & Decision Tree showed the second highest accuracy of 93.33%. After cross-validation, accuracy for Random Forest was 96.05% & for Decision Tree 94.97%.

## 2 Literature Review

The success of a movie is largely determined by how it has been justified from various viewpoints. In the early days, many works have been done on this domain where they put gross box office revenue first. ([1–4]). Using IMDb data, a few earlier publications ([4–6]) have predicted the box-office gross of a film using stochastic and regression models. A handful of them used the revenue to determine if a movie was a success or failure and then used double classifications for forecasting. Its box office revenue does not solely determine the success of a movie. A movie's success is determined by various factors such as actors/actresses, directors, budget, release month, background story, and so on. Only a few individuals had developed a prediction model based on a few pre-released features. [7]. In most cases, only a few features are taken into account. As a consequence, their models were ineffective. They overlooked audience participation once again, even though the audience mostly determines the success of a film. Even though some peoples adopt numerous implementations of NLP for sentimental analysis ([8, 9]) and assembled movie reviews for their research purpose. However, the prediction accuracy is determined by the size of the test domain. A little domain is not good for estimation. Again, the majority of them ignored critics' reviews. Moreover, user reviews as fans of the actor/actress may be biased and fail to provide an unbiased opinion.

In their research, Latif & Afzal [14], used several features (MPAA Rating, Awards, Screenplays, Opening Weekend, Meta-score, and Budgets) to classify movies into four categories based on their IMDB user ratings: Terrible, Poor, Average, and Excellent. The authors used a variety of classifiers, including logistic regression, simple logistic, multilayer perceptron, J48, naive Bayes, and PART, to achieve a classification accuracy of 84.15%, 84.5%, 79.07%, 82.42%, 79.66%, and 79.52% respectively. But their data was inconsistent and highly noisy, as stated in their paper. For that, they have used Central Tendency to fill missing values for different attributes.

Lash, and Zhao's [10] main contribution was designing a decision support system based on machine learning techniques, social network analysis, and text mining. They get information from a variety of sources. A few examples are Twitter, YouTube comments, blogs, news stories, and movie reviews. They looked at motion picture success from three perspectives: audience, release, and movie. They used this system to predict movie productivity, not revenue. The authors used linear logistic regression to classify movies

in order to forecast their profitability, which has a 77.1% accuracy rate. BoxOfficeMojo and IMDb were used to collect their original data. They only looked at movies released in the United States and left out any international films from their research.

To forecast a movie's box-office success, Mr. Sivasantoshreddy, Mr. Kasat, and Mr. Jain utilized a methodology called hype analysis [11]. For hype analysis, they mostly used Twitter data. The primary premise of hype analysis is that a film's success is mostly determined by its very first weekend earnings and the amount of hype it generates before its release. Firstly, they tried to find the total number of tweets. Here are a few factors to consider when measuring hype. The first step is determining the "number of relevant tweets per second." The second factor is "Find the number of posted tweets by different users." The third factor is to calculate "the reach of a particular tweet." The phrase "reach of a tweet" refers to the fact that the value of various people's tweets varies. They calculated the reach of a tweet by counting the number of followers of a certain user. They computed the number of linked tweets every second, and the second element is to "Find and Calculate a Tweet's Reach. In their investigation, they used the hype factor, the number of screens, and the average ticket price per show. No language processing technique was used to assess whether the tweet was good or negative. Before a movie was premiered in cinemas, a neural network was used to predict its financial success. [12]. This forecasting challenge was transformed into a nine-class classification task. The model was portrayed with only a few features.

Using Lydia's quantitative news data, Zhang and Skiena attempted to improve gross movie forecast using News analysis [13]. There are two models available (Regression and KNN models). However, they have only considered movies with a high budget. If a common word was chosen as a name, the model would fail, and it would be unable to forecast if there was no news about a movie.

Few researchers applied sentimental analysis to the social network to make their forecast [15]. Their study was focused on an intensity and positivity analysis of IMDb's Oscar Buzz sub-forum. Movie critics have been considered as their predictive perspective. When some words were used for negative meaning, however, the model produced an inaccurate result. In certain circumstances, neural network analysis was used to predict the success of a movie ([7, 18]).

Based on social media, social networks, and hype analysis, several studies calculated positivity and the number of comments associated with a certain movie ([16, 17, 19, 20]). In addition, few individuals predicted movie box office success based on Twitter tweets and YouTube comments. The forecast's accuracy will be uncertain in both cases, and the result will be unsatisfying. A limited domain is not a good concept for measuring. The majority of prior studies focused on features that were accessible either before or after the release of a film. Despite the fact that some studies considered both sorts of features, only a few were counted in that case. The probability of better prediction accuracy increases if the number of features increases.

### 3 Data Description and Methodology

In this chapter, we discussed the workflow, data extraction, cleaning, preprocessing, feature extraction, and several machine learning techniques we have used in our research.

### 3.1 Data Acquisition

Eight hundred thirty-eight films in our dataset were released between 2006 and 2018. We used the movies till 2018 because some fields of data were missing in latest films. Our primary data sources are Metacritic, IMDb, Box Office Mojo, and Rotten Tomatoes. The IMDbPy library in Python is used to collect IMDb ratings, IMDb votes, genre, directors, and casts. IMDb doesn't provide business data. We took movie budget value from Box Office Mojo and The Numbers. Another helpful aspect is the Metacritic rating, which may be found on the Metacritic website. In this dataset, we used two types of reviews. We used the Tomato rating from Rotten Tomato and the Audience Rating.

### 3.2 Data Preprocessing

Initially, we had 1764 films in our dataset. Then we recognized that several movies had no available data. The main issue was that most of the movies were missing budget data. First, we searched IMDB for the budget. IMDb does not have a budget for all films. Then we looked for a budget on Wikipedia, Box Office Mojo, Rotten Tomatoes, and The Numbers. We gathered budgets from Box-Office Mojo for some movies and The Numbers for others. We deleted any movie with an unknown year in addition to all the TV series and short movies (running time less than 60 min). After deleting the movies that did not have all of the information available, we finally had our dataset of 838 movies. The summary of our dataset is shown in Table 1.

A critical phase of data preprocessing is feature scaling. This approach normalizes the range of independent variables or data features. The algorithms can train quickly and avoid being stuck in local optima with the aid of feature scaling. In doing so, any data leaks throughout the model testing process would be avoided. We have scaled the data using machine learning techniques like logistic regression, neural networks, and others that usually utilize gradient descent as an optimization technique. Our chosen distance methods, KNN and SVM, are mainly influenced by the range of features. The scaling strategy of standardization has been applied. As a result, the attribute's mean is reduced to zero, and the distribution that results has a unit standard deviation (Fig. 1).

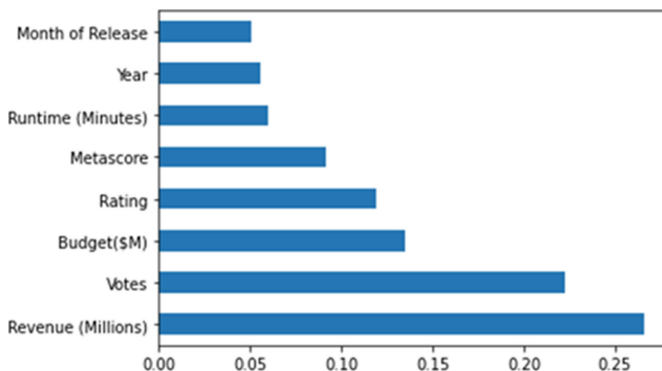


Fig. 1. Feature scaling

Missing information is a crucial component of the procedure. It's a serious problem for data analysis since it misrepresents the findings. Some characteristics in our dataset weren't accessible for all movies. As a result, we had to deal with the empty cells that were considered to be NaN. We used common-point data imputation methods to cope with missing data. In common-point imputation, the middle point, or the most frequently selected value, is used. Since this approach has three different middle values—mean, median, and mode—that are suitable for numerical data, we used mean values.

### 3.3 Feature Extraction

A dataset's features can be used to estimate a movie's rate of success. Most of the models taken into account in earlier research have a fairly small number of features. The majority of the features from the internet were included in this paper. Pre-released features and post-release features were the two distinct categories of features that we examined in this study. Pre-released features are used to assess the probability that upcoming movies will be a success. Pre-released features include the movie's budget, the month it will be released, the star power of the actors and actresses, and the star power of the director. After a few weeks after the movie's release, post-release features are helpful for increasing forecast accuracy. The post-released features in this study are the IMDb rating, IMDb votes, Meta Score, Run Time, and Gross Revenue.

Among all features genre plays an important role. Different audience have different choices. Some are fond of thriller; some are fond of action. Figure 2 shows the number of movies against each genre.

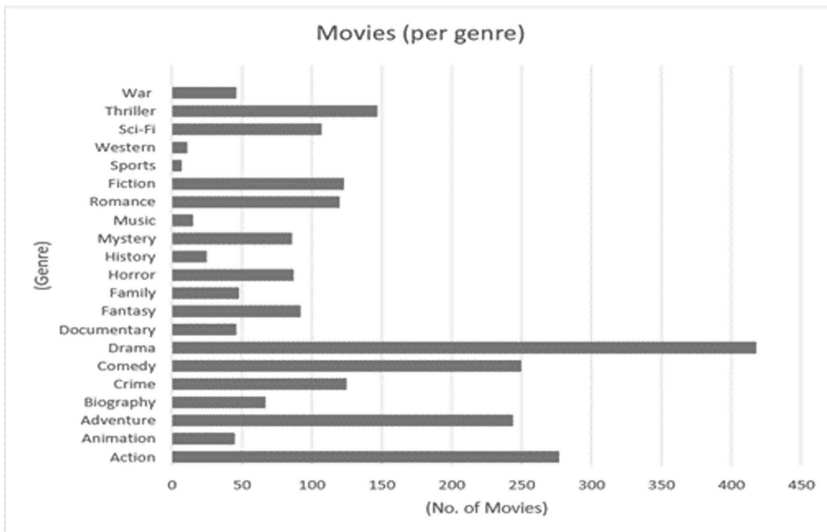
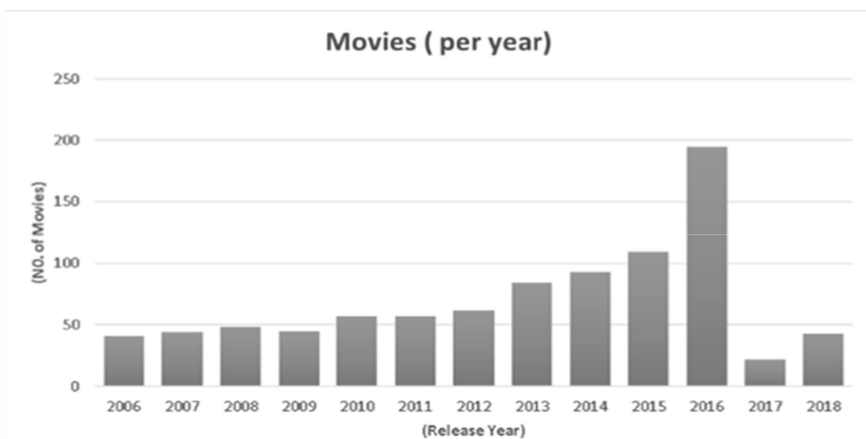


Fig. 2. Movies per Genre

**Table 1.** Dataset Summary

Features	Types	Mean	Min	Max	Std. Dev	Sources
Runtime (minutes)	Integer	114.661888	66.000000	187.000000	18.469466	IMDb
Rating	Float	6.816249	1.900000	9.000000	0.876500	IMDb
Votes	Integer	1.934572e + 05	1.780000e + 02	1.791916e + 06	1.931026e + 05	IMDb
Revenue (\$)	Float	84.970550	0.010000	936.630000	104.857058	IMDb, BoxOfficeMojo, The Numbers
Budget (\$)	Float	60.357918	0.007000	300.000000	61.860066	IMDb, BoxOfficeMojo, Wikipedia
Metascore	Integer	59.698925	11.000000	100.000000	16.949691	Meta Critic
Release Month (number)	Integer	6.672640	1.000000	12.000000	3.428753	IMDb
Cast Star Power	Integer	6.462313	3.500000	8.200000	4.986323	IMDb
Director Star Power	Integer	6.736406	4.300000	8.400000	5.334120	IMDb

We included the year as a part of the features in order to differentiate between movies having the same title but different years of release. Figure 3 shows the movies released in each year.

**Fig. 3.** Movies per Year

Film budgeting is, certainly one of the most essential aspects of the production process. Throughout the life cycle of a film, the budget plays a crucial role.

Another key aspect in the motion picture industry is the release date. For example, releasing a film in theaters at the beginning of summer permits the film to earn cash throughout the entire summer, which has a bigger demand than the fall season. Figure 4 shows the movies month of release.

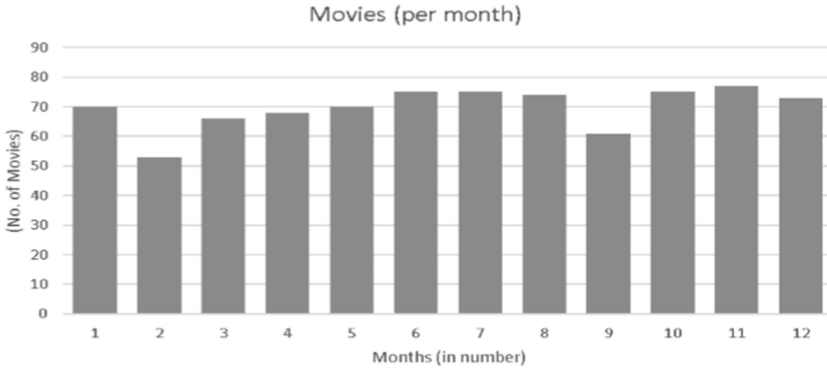


Fig. 4. Movies per Month.

Most often revenue is the key to predicting whether a movie is succeed or not. For the creative industries, forecasting and analyzing these earnings is important, and it’s often a source of interest for fans.

We considered audience votes and score from the websites like IMDb and Meta-critic’s. The number of votes reflects the number of individuals who enjoyed that particular film.

It’s crucial that a film tells its story without being either too short or too long to be enjoyable. A comedy, for example, should never last longer than 110 min. Action/drama film with strong characters and a compelling story should be between 110 and 135 min.

### 3.4 Data Transformation

Table 2 shows the third phase, data integration, and transformation where we classify our target class into five classes. Rather than giving only two output “flop” or “blockbuster” [3], we make five classifications ranging from “Flop” to “Blockbuster”.

Similarly, we have classified our other features such as budget, IMDb rating, Metascore, votes, cast star power, director-star power into different classes to predict our target model.



**Table 2.** Target class classification

Class	Range (\$)
A	Revenue $\leq$ 0.5M (Flop)
B	0.5M < Revenue $\leq$ 1Million
C	1M < Revenue $\leq$ 40Million
D	40M < Revenue $\leq$ 150Million
E	Revenue > 150 Million (Blockbuster)

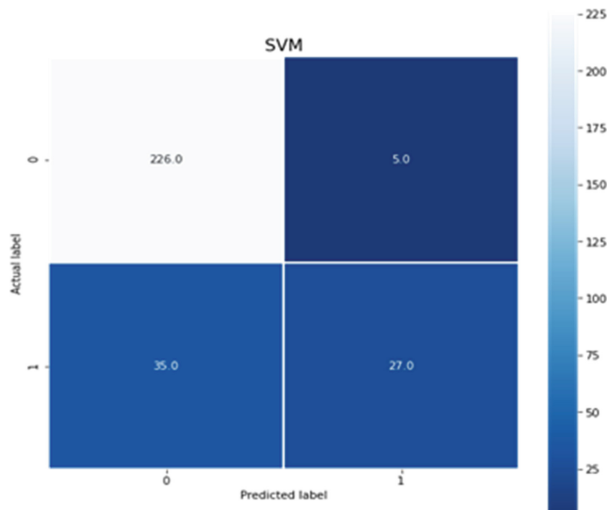
## 4 Result and Analysis

In this chapter, we present the result of our experiment for movie success prediction and discussed the results of different machine learning methodologies. From this study the prediction of movie success is identified.

### 4.1 Support Vector Machine

We have used SVM for all the pre-release and post-release features in our dataset. And it produced an accuracy of 93.25%, which is shown in the figure with the confusion matrix.

We also applied K-fold cross-validation. The mean accuracy of K-fold cross-validation using SVM is 93.40% (Fig. 5).

**Fig. 5.** Confusion Matrix of SVM

### 4.2 Random Forest

Again we used Random Forest for all the features in our dataset. And it produced an accuracy of 94.76%, which is shown in the figure with the confusion matrix. Random Forest is also used to perform K-fold cross-validation. The results of K-fold cross-validation with Random Forest with a mean accuracy of 96.05% (Fig. 6).

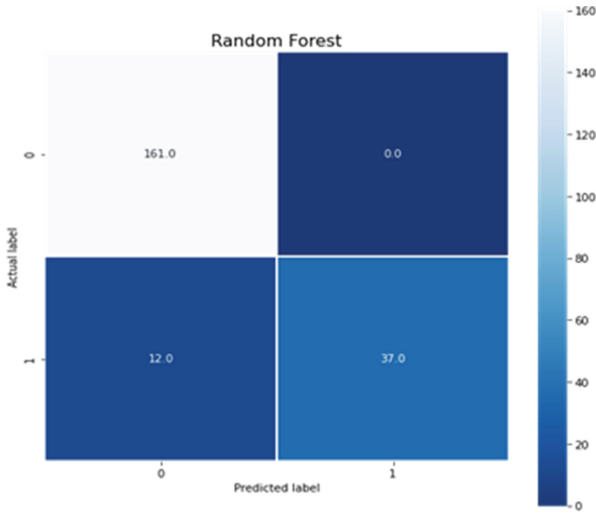


Fig. 6. Confusion Matrix of Random Forest

### 4.3 K-Nearest Neighbor

Used Random Forest for all the pre-release and post-release features in our dataset. And it produced an accuracy of 88.57%, which is shown in the figure with the confusion matrix. The mean accuracy of K-fold cross-validation using KNN is 88.66% (Fig. 7).

### 4.4 Naïve Bayes

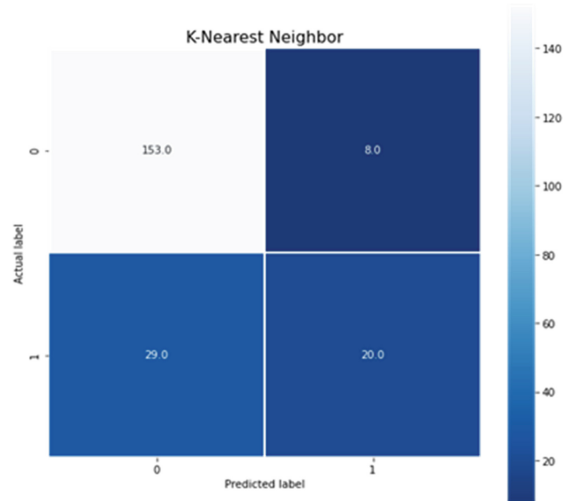
We implemented Naïve Bayes for all the pre-release and post-release features in our dataset. And it produced an accuracy of 88.67%, which is shown in the figure with the confusion matrix (Fig. 8).

We also applied K-fold cross-validation for Naïve Bayes. The mean accuracy of K-fold cross-validation using Naïve Bayes is 89.60%.

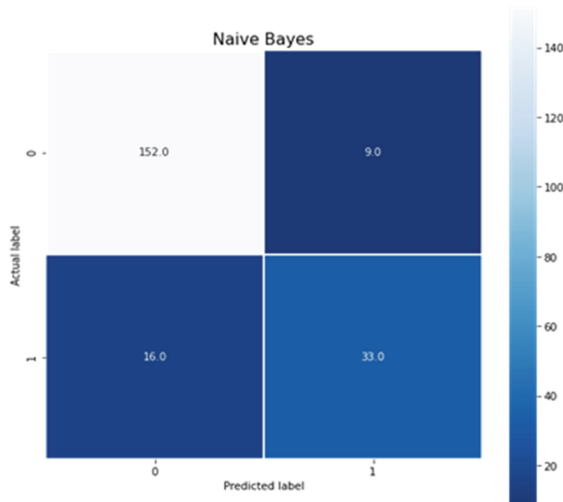
### 4.5 Logistic Regression

We implemented Logistic Regression for all the pre-release and post-release features in our dataset. And it produced an accuracy of 89.04%.

We also applied K-fold cross-validation for Naïve Bayes. The mean accuracy of K-fold cross-validation using Naïve Bayes is 91.15% (Fig. 9).



**Fig. 7.** Confusion matrix of KNN

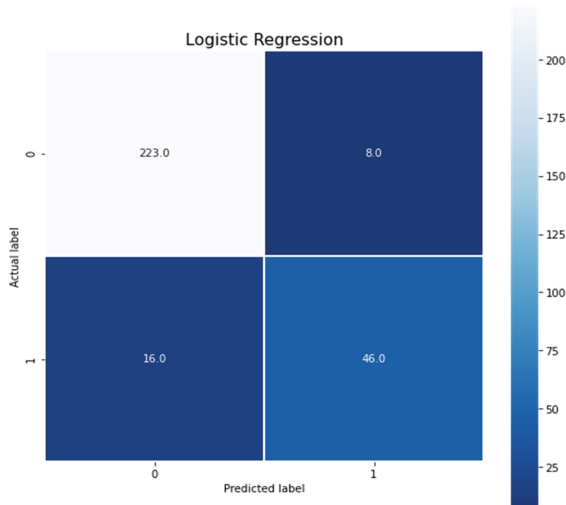


**Fig. 8.** Confusion matrix of Naïve Bayes

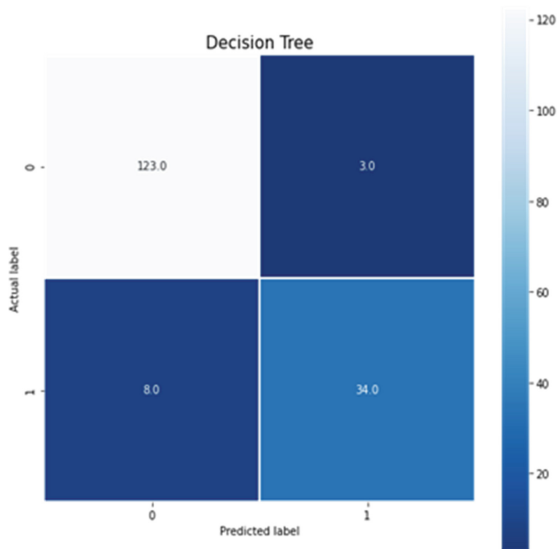
## 4.6 Decision Tree

We applied Decision Tree for all the pre-release and post-release features in our dataset. And it produced an accuracy of 93.33%, which is shown in Fig. 4.8 with the confusion matrix.

With Decision Tree, we performed K-fold cross-validation. K-fold cross-validation with Logistic Regression for each fold, with a mean accuracy of 94.97% (Fig. 10).



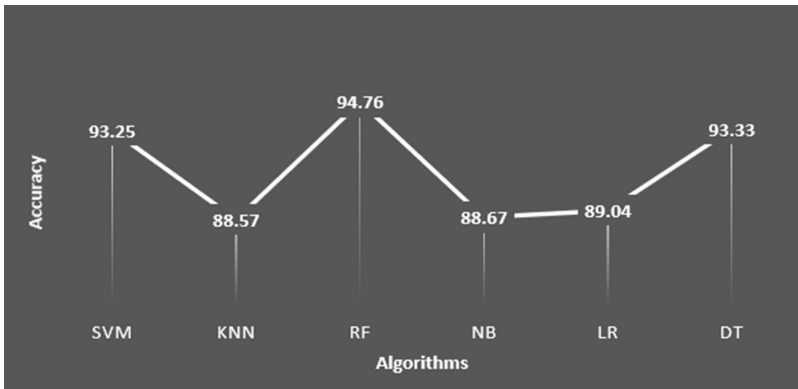
**Fig. 9.** Confusion Matrix of LR



**Fig. 10.** Confusion Matrix of DT

#### 4.7 Performance Analysis

A performance comparison among machine learning algorithms has been illustrated. It shows the accuracy of each algorithm used in your research. Some performed well and some showed less performance compared to others. We got accuracy for SVM: 93.25%, Random Forest: 94.76%, KNN: 88.57%, Naïve Bayes: 88.67%, Logistic Regression: 89.04% and Decision Tree: 93.33% (Fig. 11).



**Fig. 11.** Performance Analysis of Algorithms

After applying Cross Validation with each of algorithm, the performance increased surprisingly. Here K-fold cross-validation obtained accuracy of SVM: 93.40%, KNN: 88.66%, RF: 96.05%, NB: 89.6%, LR: 91.15% and DT: 94.97% which impacted our model significantly.

## 5 Conclusion and Future Work

The goal of this study was to develop a method for predicting movie success based on prior research. It wasn't a complete success, but it does demonstrate that other features are necessary for an accurate forecast. The number of theaters, the MPAA rating, or a film's sequel were not included in our assessment of features. It's difficult to predict a movie's sequel; some movies have generated a lot of money solely from their previous sequels. In a few other works, the sequel is likewise neglected [10]. Some research examined just pre-release features for prediction ([10, 12]), whereas others primarily examined post-release data ([1, 2]). However, while generating predictions in our research, we considered both features. In comparison to other methods, Random Forest produced the best accuracy of 94.76%. After cross-validation, Random forest once more has the highest accuracy of 96.05%. According to our research, some features have a bigger impact on a movie's success than others. Particularly affected were the budget and revenue. Furthermore, we have demonstrated that IMDB rating may be a reliable indicator of movie success.

Some aspects of the present research work can be further investigated and improved. Based on the literature reviews and studies conducted in this thesis, the following recommendations are proposed:

First, as additional features have an impact on performance rate, we intend to raise the number of features for future investigation. Second, we'll utilize more data for training, which should enhance the model's performance, and we'll use neural networks to increase performance.

## References

1. Apala, K.R., Jose, M., Motnam, S., Chan, C.-C., Liszka, K.J., de Gregorio, F.: Prediction of movies box office performance using social media. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Niagara Ontario Canada, pp. 1209–1214 (2014). <https://doi.org/10.1145/2492517.2500232>
2. Gopinath, S., Chintagunta, P.K., Venkataraman, S.: Blogs, advertising, and local-market movie box office performance. *Manag. Sci.* **59**(12), 26352654 (2013). <https://doi.org/10.1287/mnsc.2013.1732>
3. Mestyán, M.C.A., Yasseri, T., Kertész, J.: Early prediction of movie box office success based on wikipedia activity big data. *PLoS ONE* **8**(8), e71226 (2013)
4. Simonoff, J.S., Sparrow, I.R.: Predicting movie grosses: winners and losers, blockbusters and sleepers. *Chance* **13**(3), 1524 (2000)
5. Chen, A.: Forecasting gross revenues at the movie box office, Working paper, University of Washington, Seattle, WA (2002)
6. Sawhney, M.S., Eliashberg, J.: A parsimonious model for forecasting gross box-office revenues of motion pictures. *Mark. Sci.* **15**(2), 113131 (1996)
7. Sharda, R., Meany, E.: Forecasting gate receipts using neural network and rough sets. In: Proceedings of the International DSI Conference, p. 15 (2000)
8. Pang, B., Lee, L.: Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, p. 7986 (2002)
9. Chaovalit, P., Zhou, L.: Movie review mining: a comparison between supervised and unsupervised classification approaches. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS) (2005)
10. Lash, M.T., Zhao, K.: Early predictions of movie success: the who, what, and when of profitability. *J. Manag. Inf. Syst.* **33**(3), 874903 (2016)
11. Sivasantoshreddy, A., Kasat, P., Jain, A.: Box-office opening prediction of movies based on hype analysis through data mining. *Int. J. Comput. Appl.* **56**(1), 15 (2012)
12. Sharda, R., Delen, D.: Predicting box-office success of motion pictures with neural networks. *Expert Syst. Appl.* **30**(2), 243–254 (2006)
13. Zhang, W., Skiena, S.: Improving movie gross prediction through news analysis. In: 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (2009)
14. Latif, M.H., Afzal, H.: Prediction of movies popularity using machine learning techniques. National University of Sciences and Technology, H-12, ISB, Pakistan (2016)
15. Jonas, K., Stefan, N., Daniel, S., Kai, F.: Predicting movie success and academi awards through sentiment and social network analysis. University of Cologne, Pohligrasse 1, Cologne, Germany (2008)
16. Duan, J., Ding, X., Liu, T.: A gaussian copula regression model for movie box-office revenue prediction with social media. In: Zhang, X., Sun, M., Wang, Z., Huang, X. (eds.) CNCSMP 2015. CCIS, vol. 568, pp. 28–37. Springer, Singapore (2015). [https://doi.org/10.1007/978-981-10-0080-5\\_3](https://doi.org/10.1007/978-981-10-0080-5_3)
17. Doshi, L., Krauss, J., Nann, S., Gloor, P.: Predicting movie prices through dynamic social network analysis. *Procedia. Soc. Behav. Sci.* **2**(4), 64236433 (2010)
18. Rhee, T.G., Zulkernine, F.: Predicting movie box office profitability: a neural network approach. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) (2016)
19. Liu, T., Ding, X., Chen, Y., Chen, H., Guo, M.: Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools Appl.* **75**(3), 1509–1528 (2014). <https://doi.org/10.1007/s11042-014-2270-1>

20. Zhang, Z., Li, B., Deng, Z., Chai, J., Wang, Y., An, M.: Research on movie box office forecasting based on internet data. In: 2015 8th International Symposium on Computational Intelligence and Design (ISCID) (2015)
21. Valenti, J.: Motion Pictures and Their Impact on Society in the Year 2000, speech given at the Midwest Research Institute, Kansas City, 25 April 1978, p. 7 (1978)
22. Litman, B.R., Kohl, L.S.: Predicting financial success of motion pictures: the 80s experience. *J. Media Econ.* **2**(2), 3550 (1989)
23. Hayward, S.: *Genre/Sub-genre in Cinema Studies: The Key Concepts*, 3rd edn., pp. 185–192. Routledge, Abingdon (2006)