

DOI: https://doi.org/10.48009/3_iis_2024_129

Predicting popularity: Machine learning insights into movie team patterns and online ratings

Qiunan Zhang, *Alabama State University, qiunan.zhang1988@gmail.com*

Yuelin Zhu, *University of Memphis, yzhu6@memphis.edu*

Xihui Zhang, *University of North Alabama, xzhang6@una.edu*

Mark G. Terwilliger, *University of North Alabama, mterwilliger@una.edu*

Abstract

According to the 2021 Film and Video Global Market Report, the global film and video market is expected to reach \$318.2 billion by 2025 and \$410.6 billion by 2030. The financial success of a movie, though, is largely uncertain due to many pertinent factors. Understanding which factors influence the financial success of a movie is crucial for the industry. The online movie rating is often used to measure a movie's success and has been investigated in many IS studies. The factors influencing online movie ratings can be classified into two major types: internal and external. Internal factors, such as actors, actresses, directors, and "stars," influence movie ratings and are related to the movie cast and crew. External factors, such as economics, motivations for moviegoing, and advertising effects, are movie rating influencers from the outside environment. Few studies focus on both the effect of individual and team characteristics on online movie ratings. Our research question is: How do team patterns, such as actors/actresses, experienced crew, and collaboration among team members, influence movie ratings? Using both decision tree model and neural network machine learning (ML) techniques, this research investigates the effects of movie team patterns on movie ratings. We make comparisons using OLS (ordinary least squares) regression for each ML technique. The results indicate that some crews and casts of a movie, especially experienced directors, play a critical role in determining the online movie rating. The results also demonstrate the power of the decision tree model and neural network in predicting online movie ratings.

Keyword: online movie ratings, team pattern, machine learning, decision tree, neural network

Introduction

The movie industry is a multi-billion-dollar business (Lash & Zhao, 2016). According to the 2021 Film and Video Global Market Report, the global film and video market reached nearly \$234.9 billion in 2020. The market is expected to reach \$318.2 billion by 2025 and \$410.6 by 2030. The financial success of a movie, though, is largely uncertain. Understanding which factors influence the financial success of a movie is crucial for academics and the movie industry. Many researchers have undertaken the task of predicting movie success using various approaches (Lash & Zhao, 2016). Movie quality is always used to measure the success of a movie. Many factors usually influence movie quality, but all factors can be classified into two types: internal and external. Internal factors, such as actors, actresses (e.g., Wallace et al., 1993), directors (e.g., Kindem, 1982), and "stars" (e.g., Ravid, 1999), influence movie quality from the movie cast and crew. External factors, such as economics, the motivations for moviegoing, and the advertising effects (e.g., Austin, 1989), are those influencers of movie quality from the outside environment. Many studies have proven that movie quality and word-of-mouth directly and significantly affect the financial success of the

movie, such as box office sales and other related movie products (Wallace et al., 1993). As a combination of both art and creative works, the individual effect plays a deterministic role that could influence the movie's quality. Lash and Zhao (2016) also showed that team works plays a critical role in a movie achieving success.

Online movie ratings can lead to the movie's financial success, encouraging people to go to the cinema (e.g., Zhou et al., 2019), watch the movie online (e.g., Feng, 2017), and buy DVDs of the movie. Due to under-reporting bias, online ratings and reviews may not represent the general consensus options (Hu et al., 2006; Koh et al., 2010). The online movie rating is measured by its quality to some extent (Koh et al., 2010), as well as other aspects, such as stars (e.g., Feng, 2019). Few studies, however, focus on both the effect of individual and team characteristics on movie ratings.

Recently, machine learning technologies, such as decision trees and neural networks, have been adopted in many studies and proven as one of the most potent methodologies in scientific research (Xu et al., 2022). Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data. Compared to traditional statistical methods, machine learning does not necessarily require a strict and comprehensive theoretical building. Through the whole learning process, the model will be generated. These techniques could save time and are very good at handling big data (Ij, 2018).

Our research investigates the effects of team patterns on online movie ratings. Our research question is: How do team patterns influence online movie ratings? The main objective of this research is to investigate the effect of team features, such as stars, previous members' experiences, and collaboration, on movie ratings. A model is established to predict online movie ratings based on team patterns by applying data mining and machine learning techniques. Hypotheses are developed in the following sections through a literature review. Based on the hypotheses, the methodology is determined and presented. We share the results of our research following the methodology section. The paper concludes with a discussion of its contributions, limitations, and future research.

Literature review and theoretical development

Based on our research question, three streams of literature are reviewed. One stream is the individual effects on the success of a movie. Wallace (1993) showed that the power of stars has positive effects on the success of a movie. The skills and knowledge of stars could improve movie quality. Other research certifies that crews of movie teams, such as directors and writers, could determine the success of a movie. The director's previous experiences and expertise positively influence the success of a movie as well (Lash & Zhao, 2016; Lutter, 2014). Recent studies also demonstrated that directors, stars, and other team members could play a critical role in determining the success of a movie (e.g., Memon et al., 2024).

A second literature stream is teamwork and team performance. Team patterns and diversity could help bring creative ideas to creative activity, and such "team chemistry" could eventually improve team performance (Amiri et al., 2024; Lash & Zhao, 2016). Also, the study of Long et al. (2014) shows that repeated collaboration is beneficial for team performance. Through repeated collaboration in the same team, team members will get to know each other, build the mutual trust, and improve knowledge sharing and the atmosphere among team members. Consequently, the performance and outcome of the team will be improved.

The third stream is the characteristics of online reviews and ratings. Online users prefer to express their opinions and ideas, including ratings based on their subjective perceptions (e.g., Koh et al., 2010; Zhang et

al., 2021). The public opinion of social media on online movie ratings also plays a critical role in determine movie rating, such as hashtags of X (Twitter) (Bhadrashetty & Patil, 2024). The effect of famous actresses/actors and directors may exceed the effects of their own capabilities, which contributes to the movie quality of online movie ratings (Feng, 2019). Also, some fixed combinations of actresses and actors may impact online ratings due to the preferences of users (coupling culture) (Li, 2022).

While the existing literature provides insights into the individual effects and teamwork dynamics in movie production, a comprehensive analysis with most recent data mining techniques, such as machine learning, to predict movie success is lacking. Specifically, there is a gap in understanding how the interplay among star power, team collaboration, and other characteristics of movie (e.g., genre), collectively influences movie success.

Through literature review, it is not hard to understand or imagine the power of stars, experienced directors, and collaboration among movie teams from the literature. Based on our research question, research objective, and existing literature, three hypotheses are developed:

- **H1:** “Star” has a positive effect on movie rating; that is, a movie with more stars has a higher online movie rating.
- **H2:** The experienced crew positively affects movie rating; that is, a movie with a more experienced crew has a higher online movie rating.
- **H3:** Repeated collaboration positively affects movie rating; that is, a movie with a team with more collaboration times in the past has a higher online movie rating.

Methodology

Based on our hypotheses, Ordinary Least Squares (OLS) regression will be used in the first step. OLS regression is a common technique to investigate the relationship between one or more independent quantitative variables and a dependent variable (Noreen, 1988). OLS regression could provide the minimum squares error (SSE) in estimation, which is easy to implement. The OLS regression will be applied to test the hypotheses.

The results of OLS are not straightforward, however, to all people and need to be interpreted by professionals. The coefficient for variables only reflects the trend and relationship but cannot explain which actions should be taken precisely. The decision tree model will be an excellent tool to show which features of the movie team will lead to higher online movie ratings and help movie investors make clear decisions. The decision tree is a supervised learning and classification technique that can model data with nonlinear relationships between variables and handle interactions between variables. The decision tree is a flowchart-like tree structure. In the structure, the internal nodes denote tests on attributes, the branches represent outcomes of tests, and the leaf nodes represent class labels or a class distribution. There are several algorithms commonly used, such as Hunt’s Algorithm, CART, ID3, C4.5, SLIQ, SPRINT. These algorithms are non-parametric and can efficiently deal with large, complicated datasets without imposing a complicated parametric structure (Song & Lu, 2015). For the decision tree model development, creating nodes with a homogeneous class distribution is preferred. Generally, Entropy, Gini, or Classification error is used as the criterion for creating nodes.

Besides decision trees, another machine learning technique – an artificial neural network - will be used to set up a model to predict online movie rating scores based on team patterns and other features. Neural networks are a series of algorithms that mimic the operations of an animal brain to recognize relationships

between vast amounts of data. As such, they tend to resemble the connections of neurons and synapses found in the brain. In neural networks, information processing occurs at many identical and simple processing elements called neurons (also called units, cells, or nodes). Interneuron connection strengths known as synaptic weights are used to store knowledge (Dongare et al., 2012). Neural networks are often used as a powerful tool to cluster and classify. They group unlabeled data according to similarities among the example inputs, and they classify data when they have a labeled dataset to train on.

Data and variables

The secondary datasets of online movies and ratings were obtained from Kaggle (<https://www.kaggle.com/rounakbanik/the-movies-dataset?select=ratings.csv>). Movie information and online movie ratings are stored in different datasets. The dataset of movie information contains a large amount of JSON format data, and Excel or other applications cannot parse the JSON format data. Thus, the information on casts, crews, and genres is pulled randomly from the dataset. This task was very time-consuming. In the end, the information of movies is merged with online movie ratings based on movie ID.

Based on the research's objectives and hypotheses, the dependent variable/target and the independent variables/features are set up for OLS regression and a decision tree model, respectively. The dependent variable/target is Online movie rating. The rating score of a movie ≥ 3.5 is set as High; less than 3.5 and greater than or equal to 2.5 is set as Medium, and below 2.5 is set as Low. Besides the movie ratings data from Kaggle, the movie rating data from the IMDB database is also used as a reference.

Based on the actual situation, pulling all the stars and experienced crews from the movies is challenging since all of the information-generation processes are operated manually. The stars number of the first three actors and experienced directors is only counted in the movie. The behavior of "self-directed and acted" may influence the quality of the movie because this behavior always improves the outcome of creative activity (Kovalenk & Smirnova, 2015). Therefore, the Number of stars in a movie, Experienced director, Collaboration times, and Self-directed are regarded as independent variables/features. Because the dataset does not contain such information, the value computation of these independent variables/features is also needed to be done manually. The detailed rules of computation are shown as follows:

- Collaboration times (CT) = Max (Act 1, 2, 3 & Director's collaboration times). If the movie is their first collaboration, count as 1. If they worked together in a previous movie, count 2; if they worked together twice before, count 3.
- Self-directed (SD): if a director of one movie acts as the first three actors/actresses, count one.
- Star number (SI): if one actor/actress was one of the first three actors/actresses in a previous movie, then regard him/her as the star. Also, count how many stars are in this movie. Min is 0, and max is 3.
- Experienced director (ED): if the director of this movie was also the director of a previous film, then marked 1.

Based on the above rules, a total of 1,009 movies from the original dataset were randomly selected and processed. Python programs were used to run OLS Regression, a Decision Tree model, and a Neural Network on Google Colab (<https://colab.research.google.com/drive/1jbUHOSmd-PpWjCcLMKHZtydD3nYWozyU#scrollTo=pCzxYXfjAsw0>). Besides the regular libraries such as numpy and pandas for data analytics in Python, statsmodels was used for OLS regression, and sklearn was used for the Decision Tree model and Neural Networks in this research. The results are presented in the following section.

Results

OLS results

The result of OLS is shown in Figure 1. From the OLS model, the R square is 0.198, which means there is a 19.8% variance in movie rating that could be influenced by Collaboration times (CT), Self-directed (SD), Number of stars (SI), and Experienced director (ED). Except for Collaboration times, all the independent variables positively affect the Online movie rating. The Number of stars and Experienced director (ED) are statistically significant at a 5% significance level. The effect of an experienced director (0.3867) is more significant than the effect of stars (0.1648). Thus, **H1** and **H2** are supported, and **H3** is not supported. This result suggests that more stars and having an experienced director in a movie team will improve the movie’s online rating. It also indicates that repeated collaboration and self-directed actors may not significantly affect the online rating of a movie.

OLS Regression Results						
=====						
Dep. Variable:	Rating	R-squared:	0.198			
Model:	OLS	Adj. R-squared:	0.195			
Method:	Least Squares	F-statistic:	62.12			
Date:	Tue, 21 May 2024	Prob (F-statistic):	6.23e-47			
Time:	23:10:07	Log-Likelihood:	-871.75			
No. Observations:	1009	AIC:	1753.			
Df Residuals:	1004	BIC:	1778.			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	3.0988	0.043	71.502	0.000	3.014	3.184
CT	-0.0021	0.036	-0.058	0.954	-0.073	0.069
SD	0.0703	0.086	0.820	0.412	-0.098	0.238
SI	0.1648	0.019	8.584	0.000	0.127	0.203
ED	0.3867	0.040	9.629	0.000	0.308	0.466
=====						
Omnibus:	191.914	Durbin-Watson:	1.839			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	443.924			
Skew:	-1.035	Prob(JB):	4.01e-97			
Kurtosis:	5.505	Cond. No.	9.77			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 1: Results of OLS Regression

The OLS regression with fixed effect is also implemented as the robustness check. The results of OLS regression with a fixed effect (see Figure 2) are similar to that of OLS regression. Furthermore, the R-square of this model is 0.226, which means there is a 22.6% variance in movie rating that could be influenced by Collaboration times (CT), Self-directed (SD), Number of stars (SI), Experienced director (ED), and different Genres. The result shows that some movie genres significantly affect movie ratings statistically, such as Action, Adventure, and Animation. They all have significant effects on online movie ratings statistically. This fact also indicates that the results of the OLS regression model are robust to our hypotheses.

OLS Regression Results

```

=====
Dep. Variable:          Rating    R-squared:                0.226
Model:                  OLS      Adj. R-squared:           0.210
Method:                 Least Squares  F-statistic:              13.76
Date:                   Tue, 21 May 2024  Prob (F-statistic):       2.86e-42
Time:                   23:18:27    Log-Likelihood:           -853.79
No. Observations:      1009        AIC:                      1752.
Df Residuals:          987         BIC:                      1860.
Df Model:               21
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	2.9218	0.044	65.678	0.000	2.834	3.009
CT	0.0027	0.036	0.074	0.941	-0.068	0.074
SD	0.0736	0.086	0.857	0.391	-0.095	0.242
SI	0.1692	0.019	8.689	0.000	0.131	0.207
ED	0.3705	0.040	9.185	0.000	0.291	0.450
Genres_Action	0.1192	0.056	2.131	0.033	0.009	0.229
Genres_Adventure	0.1684	0.070	2.407	0.016	0.031	0.306
Genres_Animation	-0.3786	0.165	-2.295	0.022	-0.702	-0.055
Genres_Comedy	0.1794	0.047	3.850	0.000	0.088	0.271
Genres_Crime	0.2572	0.069	3.711	0.000	0.121	0.393
Genres_Documentary	0.2757	0.092	2.995	0.003	0.095	0.456
Genres_Drama	0.1947	0.051	3.806	0.000	0.094	0.295
Genres_Family	0.0774	0.222	0.349	0.727	-0.358	0.513
Genres_Fantasy	-0.0050	0.101	-0.049	0.961	-0.202	0.193
Genres_History	0.3958	0.121	3.280	0.001	0.159	0.633
Genres_Horror	-0.0020	0.075	-0.027	0.979	-0.149	0.145
Genres_Music	0.4510	0.126	3.570	0.000	0.169	0.699
Genres_Mystery	0.2127	0.099	2.150	0.032	0.019	0.407
Genres_Romance	0.2747	0.071	3.855	0.000	0.135	0.415
Genres_Science Fiction	0.0289	0.121	0.239	0.811	-0.208	0.265
Genres_Thriller	0.1251	0.071	1.761	0.078	-0.014	0.264
Genres_War	0.3068	0.146	2.095	0.036	0.019	0.594
Genres_Western	0.2401	0.118	2.029	0.043	0.008	0.472

```

=====
Omnibus:                177.262    Durbin-Watson:            1.842
Prob(Omnibus):          0.000    Jarque-Bera (JB):         402.188
Skew:                   -0.968    Prob(JB):                 4.63e-88
Kurtosis:               5.412    Cond. No.:                1.73e+16
=====

```

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is 1.46e-29. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Figure 2: Results of OLS Regression with Fixed Effects

In this research, the sample is split randomly into 70% as the training set and 30% as the test set. The decision tree model is implemented with the entropy criterion, the depth of the tree as 10, and the accuracy

as 0.6369. After comparing the accuracy between the training set and the test set (see Figure 3), the accuracy does not change after depth = 3. Also, in order to save space in the report, the max depth is set as 3. The final tree structure is shown in Figure 4.

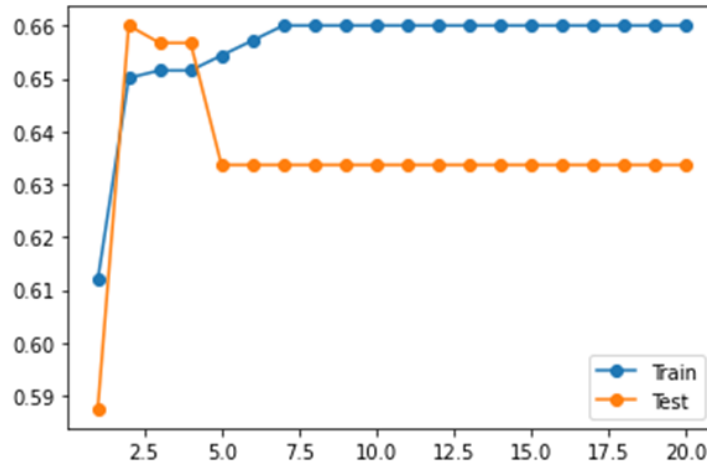


Figure 3: Accuracy of Training and Test Sets

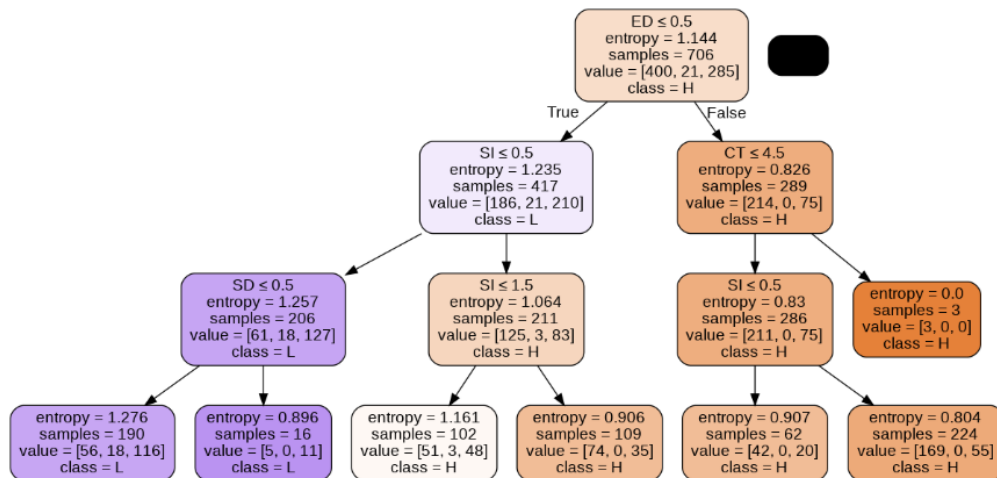


Figure 4: Decision Tree Model

The model shows that Experienced director (ED) is a vital feature leading to a high score. If a movie team has an experienced director, it may lead to a high rating score for a movie. Higher Numbers of stars (SI) also contributes to a high score for a movie. When a movie has one or more stars, the movie may have a high rating. The Self-directed and Collaboration times do not seem to influence the movie’s online rating.

A decision tree model with all Genres is also implemented (see Appendix A). The tree structure of this model is larger than the model without genres. Compared with the two confusion matrixes (see Figure 5) of these two models, there are few differences between them.

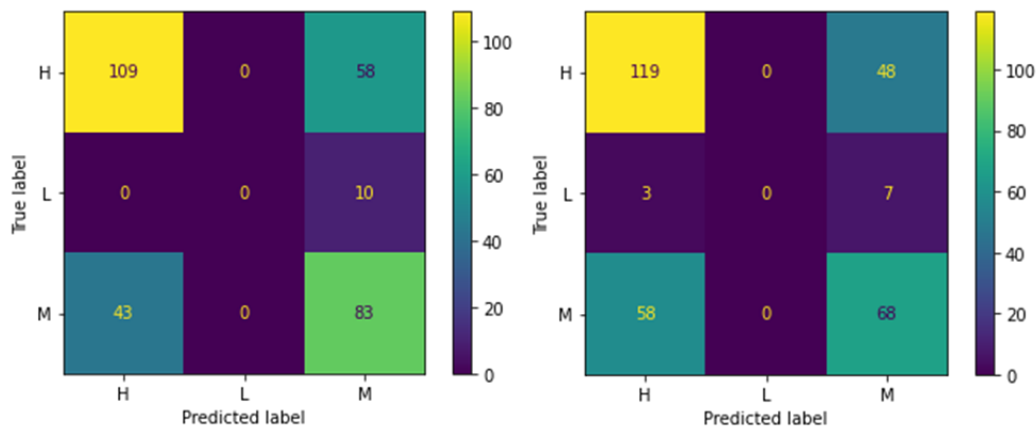


Figure 5: Confusion Matrix (Left is the model without genres, and right is the model with genres)

Based on these results, the decision tree model is reliable, and the structures of the two decision tree models are consistent with the result of the OLS regression. These findings support **H1** and **H2** strongly.

Neural networks

In the training of the neural network (NN), the rectified linear activation function (ReLU) is applied as the activation function. ReLU is a piecewise linear function that will output the input directly if it is positive; otherwise, it will output zero. Because the number of inputs for the NN is four, and the output is the rating score, which is numerical data, ReLU will be a proper function under this condition. ReLU has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance. This NN is set to have 64 nodes, and the number of epochs is set to 100. The sample data is randomly split 70% into the training set and 30% into the test set. Ten percent of the data is used as the validation set after the NN trains on the other 90% of the data in each epoch. The performance (MSE) of this NN is shown in Figure 6.

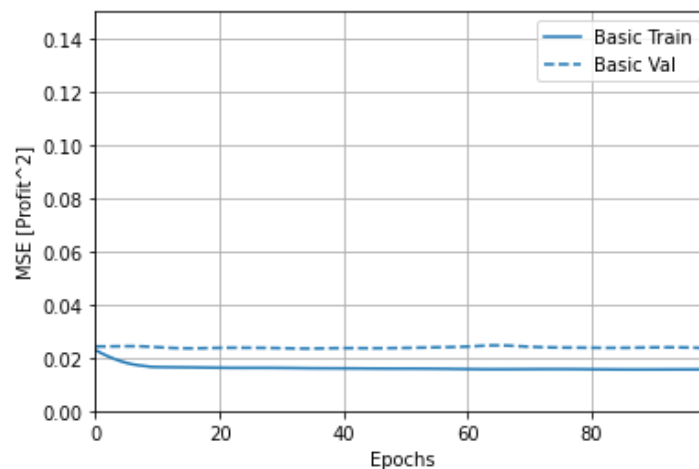


Figure 6: MSE of Neural Network Model

The MSE of this model shows that at about the 10th epoch, the MSE of the model is steady and converges to about 0.017, which means the NN has a good performance at predicting online movie rating scores based on input features.

Discussion

The OLS regression and decision tree model results show that experienced directors and stars play a critical role in influencing the online rating of the movie. The power of stars and the experienced director is strong, as expected. The experienced director is more important than the stars in determining the movie's rating. At the same time, repeated collaboration and self-directed behavior seem to have little effect on online movie ratings. The findings seem inconsistent with some past literature (e.g., Lash & Zhao, 2016; Long et al., 2014) about team works and creative activity. It is reasonable, however, in movie production. Many facts show that after the first movie's success, most movie series do not have good ratings even with the same casts and crews.

The results of OLS regression with fixed effects and a decision tree model with genres show that some genres could impact online movie ratings positively or negatively. These differences may be explained by different types of movies attracting different people, and those people will rate the movies based on their preferences. People with similar preferences may enjoy watching specific types of movies and tend to give a similar rating to a specific type of movie. Consequently, the online movie rating may be influenced by the genre of the movie.

Compared with OLS regression, the decision tree model is easier to understand and interpret. The result shows that the number of stars and experienced directors will lead to high online movie ratings. The accuracy of the decision tree model is 0.6369, which is not a high accuracy score. However, compared with the accuracy of 414 movies in the first phase, which is around 0.52, the model is improved with a larger sample size. It indicates that more movie information input may achieve a better model.

The neural network is a powerful predictive tool and has a relatively good performance. It allows multiple epochs to modify the weights to achieve higher accuracy and lower errors if adequate activation functions are selected. This research demonstrates that such a technique could predict movie rating scores well. A neural network, however, needs much work processing data (such as data scaling) and building a model. Furthermore, it may struggle with large inputs, a small sample size, or inadequate activation functions. Compared with OLS regression, it is not easy to interpret the effect of different inputs/features on outputs since the weights change from epoch to epoch until they converge. The nodes may be unclear to audiences compared to the decision tree model.

Contributions

This research has several contributions. First, the result could help the movie industry better understand how to form their team to get a higher online-rated movie and thus achieve a good ROI from the investment and a good reputation. Investors could consider the team carefully based on this result to achieve a high online movie rating and thus may achieve the financial success of a movie. For example, they may consider inviting an experienced director as their priority when the resources are limited. They could choose some specific genres of movies to achieve high-rating movies, as well.

Second, this research can enrich our understanding of the effect of team patterns on team performance in the context of movie production and provide an example of applying machine learning in such a context. This research will also enrich our understanding of online users' behaviors in the online movie rating context.

Limitations and future research

Our research also has some limitations. First of all, the sample size may not be large enough. Using 1,009 movies is not a small sample, but considering different genres and other independent variables, the sample size seems not enough to be applied with neural networks well. Second, the sample does not have enough different types of output, such as low-rating movies (see Figure 7), which may cause some bias in the model training process or poor performance of the model. Third, other influential crews, such as writers and producers, are not put into the model due to difficulties in processing data.

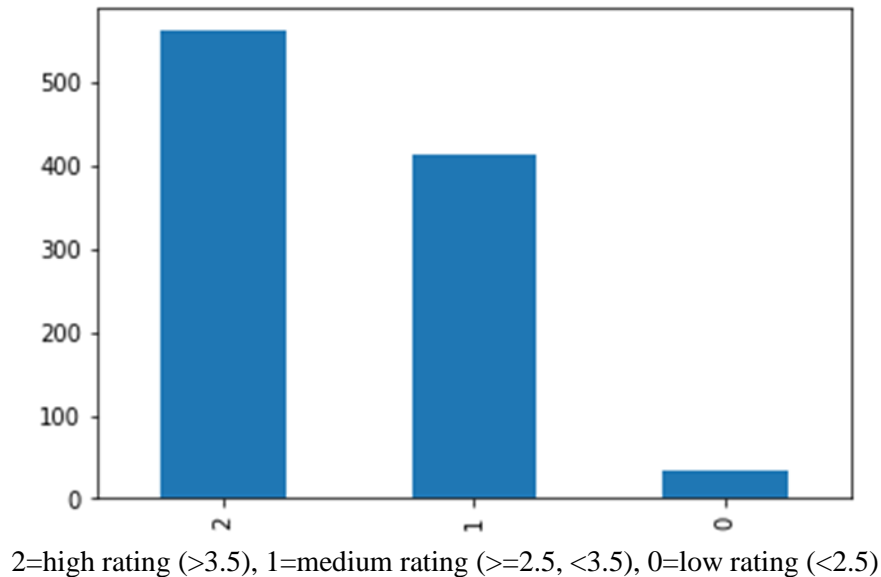


Figure 7: Movie Rating Distribution

Collaboration times may be influenced by other individual factors, such as trust in the team and the team's atmosphere influenced by individuals. Limited to the dataset, however, the relationships and factors cannot be easily interpreted. Lastly, collaboration times may not be independent of other independent variables, which may violate assumptions in the OLS regression.

Based on the limitations, our future plan is to collect more data to run the model. Other literature will be reviewed to determine how to weigh different individuals in the movie team. More crews, such as writers and producers, who are also crucial to the online movie ratings, will be integrated into our model. Other models, such as association rule mining, will be used to investigate the effects of team patterns, such as which casts and crews or which fixed combination of crews may have higher probabilities of producing high-rating movies or attracting more online users to give a high rating.

References

- Amiri, F., Sohrabi, M. R., Hannani, S., & Azadi, N. (2024). The effect of TeamSTEPPS® mobile-based training on perceptions of teamwork and team performance of operating room staff. *Perioperative Care and Operating Room Management*, 100396.
- Austin, B. A. (1989). *Immediate seating: A look at movie audiences*. Wadsworth Publishing Company.

- Bhadrashetty, A., & Patil, S. (2024). Movie success and rating prediction using data mining. *Journal of Scientific Research and Technology*, 2(1), 1-4.
- Businesswire. (2021). *Film and video global market report 2021*. Businesswire.
<https://www.businesswire.com/news/home/20210910005333/en/Global-Film-and-Video-Services-Market-Report-2021---Opportunities-and-Strategies-to-2030---ResearchAndMarkets.com>
- Dongare, A. D., Kharde, R. R., & Kachare, A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology*, 2(1), 189-194.
- Feng, G. C. (2017). The dynamics of the Chinese film industry: Factors affecting Chinese audiences' intentions to see movies. *Asia Pacific Business Review*, 23(5), 658-676.
- Feng, G. C. (2019). A comparative study of the online film ratings of US and Chinese audiences: An analytical approach based on big data. *International Communication Gazette*, 81(3), 283-302.
- Hu, N., Pavlou, P. A., & Zhang, J. (2006, June). Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In *Proceedings of the 7th ACM conference on Electronic commerce* (pp. 324-330).
- Ij, H. (2018). Statistics versus machine learning. *Nat Methods*, 15(4), 233.
- Kindem, G. (1982). Hollywood's Movie star system: A historical overview. In Kindem, G. (Eds.), *The American movie industry: The business of motion pictures* (pp.79-94). Southern Illinois University Press.
- Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, 9(5), 374-385.
- Kovlenko, N. A., & Smirnova, A. Y. (2015). Self-directed learning through creative activity of students. *Procedia - Social and Behavioral Sciences*, 166(7), 393-398.
- Lash, M. T. & Zhao, K. (2016) Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems*, 33(3), 874-903.
- Li, M. (2022). Exploration of psychological logic of CP fans in cultural consumption. *Management Science and Research*, 11(4), 102-105.
- Lutter, M (2014, July 9). *Creative success and network embeddedness: Explaining critical recognition of film directors in Hollywood, 1900-2010*. SSRN: MPIfG Discussion Paper, 14/11.
<https://ssrn.com/abstract=2464150> or <http://dx.doi.org/10.2139/ssrn.2464150>
- Memon, Z. A., & Hussain, S. M. (2024). Predicting movie success based on pre-released features. *Multimedia Tools and Applications*, 83(7), 20975-20996.
- Noreen, E. (1988). An Empirical Comparison of Probit and OLS Regression Hypothesis Tests. *Journal of Accounting Research*, 26(1), 119-133. <https://doi.org/10.2307/2491116>

- Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. *The Journal of Business*, 72(4), 463-492.
- Song, Y. Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130–135.
<https://doi.org/10.11919/j.issn.1002-0829.215044>
- Wallace, W. T., Seigerman, A., & Holbrook, M. B. (1993). The role of actors and actresses in the success of films: How much is a movie star worth? *Journal of Cultural Economics*, 17(1), 1-27.
- Long, J., Zang, Z., & Chen, C. (2014). The benefits of repeated collaboration for team performance in a Chinese context. *Social Behavior and Personality*, 42(9), 1507-1520.
- Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4), 100179.
- Zhang, Q., Onita, C. G., Ray, D. A., Banks, M. S., & Zhang, X. (2021). Understanding antecedents of online gamers' negative emotions during the COVID-19 pandemic. *Issues in Information Systems*, 22(4), 21-32.
- Zhou, Y., Zhang, L., & Yi, Z. (2019). Predicting movie box-office revenues using deep neural networks. *Neural Computing and Applications*, 31, 1855–1865. <https://doi.org/10.1007/s00521-017-3162-x>

Appendix A: Decision Tree Model With Genres

