Peertechz

MATHEMATICS AND PHYSICS GROUP

Annals of Mathematics and Physics — OPEN ACCESS JOURNAL

submitpaper@peertechz.com

ISSN: 2689-7636 DOI: https://dx.doi.org/10.17352/amp

Special Issue entitled: Application of Mathematical Methods in Economics and Finance

Research Article

Explanatory and predictive analytics for movie production efficiency by online word-of-mouth

Sangjae Lee[1]* and Joon Yeon Choeh[2]

[1]College of Business Administration, Sejong University, Seoul 05006, Korea

[2]Department of Software, Sejong University, Seoul 05006, Korea

Received: 25 September, 2023
Accepted: 12 October, 2023
Published: 13 October, 2023

***Corresponding author:** Sangjae Lee, College of Business Administration, Sejong University, Seoul 05006, Korea, Tel: +82 2 3408 3980;
E-mail: sangjae@sejong.ac.kr

Keywords: eWOM; Movie resources powers; Production efficiency; Prediction of production efficiency; Explanatory analytics for movie production efficiency; Predictive analytics for movie production efficiency

Abstract

It has become increasingly important to consider the efficiency of movies in creating box revenue while using fewer movie resources. Further, there is a lack of eWOM (online-word-of-mouth) studies regarding using the production efficiency of movies as a dependent outcome measure replacing box revenue. This study shows that production efficiency can be suggested by comparing movie resources powers, i.e., powers of actors, directors, distributors, and production companies, which are input for movie production, and the box office. For testing the validity of the measure of production efficiency, this study examines the effect of eWOM attributes, i.e., review depth, volume, rating, review sentiment, and helpfulness on production efficiency. Data envelopment analysis is adopted to produce the efficiency of movies. This study provides insights into a current movie study on eWOM by showing the effect of interaction between eWOM (review rating) and helpfulness on production efficiency. Further, this study purports to test the prediction power in predicting production efficiency using decision trees, neural networks, and logistic regression. These results show that k nearest neighbor and automated neural networks outperform the other machine learning methods in classifying efficient movies.

Introduction

As one of the rapidest developing global industries [1], it is important to suggest the performance of the movie industry [2], as it is critical for lowering market risk and facilitating the enlargement of a movie-related market [3].

Thus, movie revenue prediction has received much attention in order to understand how to promote the movie industry [4-8]. It is becoming important to show alternative measures of movie performance given the criticality of assessing the outcome of movie production. The factors that have an effect on a box office encompass eWOM (online word-of-mouth) variables, for example, review volume, review depth, review rating, and the sentiment of the review. It is important to understand that crucial factors affecting box-office revenue [2,9] are still affecting the alternative measure of movie performance.

Further, it is necessary to predict movie performance given that a great risk exists in obtaining the break-even point of new motion pictures due to uncertainty in the public's taste. Thus, it is crucial to suggest the predictive capability using independent factors like eWOM variables and the alternative measure of movie performance. Thus, our study offers the following motivations toward previous literature on movie performance.

First, this study intends to suggest an alternative measure of movie performance, i.e., production efficiency, and investigate how eWOM has an effect on production efficiency. Previous studies on movie performance indicate that movie budget comprises a crucial factor for box revenue [10], and as one of the important factors in movie budget, star powers greatly affect the box office [11,12]. Because previous studies on movies have not sufficiently examined production efficiency [13-15], our research focuses on production efficiency which is produced by comparing four resource powers, i.e., powers of actors, directors, distributors, and production companies, with box office to coherently show how much movie resources are consumed in the production of movies and represent the efficiency of spending movie resources in creating box office. The production efficiency can be considered to show how much resources are expensed in a more collective manner for obtaining box office. This study adopts Data Envelopment Analysis (DEA) to produce production efficiency and shows whether eWOM variables which are largely examined as factors affecting box office are still effective in affecting production efficiency.

Secondly, our study employs predictive analytics methods to examine their predictive performance in predicting production efficiency. It is important to test the predictive efficacy of data mining methods for this alternative measure of movie performance as the studies on the prediction of movie production efficiency are almost nonexistent.

The attributes of eWOM include the volume of review, review depth, review rating, and the number or percentage of positive or negative sentiments [16,17]. The effects of eWOM such as review rating and volume on sales have been widely studied [18-20]. Thus, our study intends to investigate the effect of these eWOM variables on production efficiency.

Second, this study aims to classify movies into efficient and inefficient movies and evaluates the prediction performance of machine learning methods (decision trees, k-nearest-neighbors) between efficient and inefficient movies. Whilst there exists previous research utilizing various machine learning methods for predicting box office, few studies have evaluated how the prediction performance turns out to be different between efficient and inefficient movies. This study purports to fill this gap. Specifically, this study utilizes DEA to suggest efficient and inefficient movies based on movie resources to create movie revenue. This study intends to adopt non-statistical and statistical methods to show a more balanced use of machine learning methods and adopts ensemble (bagging) methods to increase predictive performance by suggesting and averaging results across multiple random samples.

Movie production efficiency

Production efficiency is produced using DEA as a dependent factor measuring movie performance which is an alternative measure replacing movie revenue. While many studies examine the significance of explanatory eWOM variables for movie revenue, almost no study exists for using eWOM to explain production efficiency. Production efficiency is evaluated by comparing movie resources powers, i.e., powers of actors, directors, distributors, and production companies, concerning movie revenue. The budget of a movie has been considered a crucial factor for movie revenue [10], and star powers which have a large portion of the movie budget have been regarded as influencing movie revenue [11,12].

DEA is a non-parametric measurement approach for efficiency [21] that uses linear programming for producing a piecewise linear production frontier for creating movie efficiency. The efficient frontier uses all data as a reference sample with which each decision-making unit is assessed [22]. DEA need not a priori assumption regarding the underlying distribution function [23].

The applications of DEA encompass efficiency analysis of controls [24], the effect of electronic health record implementation on hospitals' productivity [25], strategic R&D portfolio management [26], evaluation of productivity for IT service industry [27], prioritization of patients (Rahimi et al., 2016), and movie efficiency [15]. In order to extend the line of research concerning the effect of eWOM on movie revenue, this study purports to examine the effect of eWOM on production efficiency.

Research model

Resource powers like celebrity powers should be considered important as they determine the movie budget and subsequently influence movie performance [10-12]. In order to more effectively investigate the effect of eWOM on movie performance, this study adopts production efficiency in converting movie resources powers, i.e., power (past total revenue) of actors, directors, distributors, and production companies, into the box office.

The research model is depicted in Figure 1. This model shows the impact of volume (measured as review depth and review volume), valence (measured as review rating), sentiment (measured as the percent of emotional reviews), and helpfulness (measured as review helpfulness) on the production efficiency of the movie.

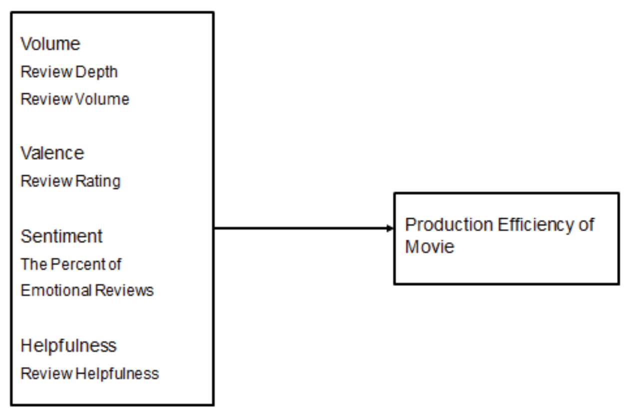

Figure 1: Research Model.

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015.
DOI: https://dx.doi.org/10.17352/amp.S1.000002

This study adopts the eWOM factors which show quantity, valence and sentiment, and helpfulness aspects of eWOM, i.e., review depth, review volume, review rating, the number of positive reviews, and helpfulness. Review depth and volume indicate the quantity aspect of eWOM facilitating awareness effect and show the extent of online review activities. As the quantity of eWOM increases promoting consumer awareness [28-30], this eWOM's awareness effect facilitates movie performance due to its positive influence on movie recognition. The dispersion of eWOM enhances the popularity of movies to build buzz, which subsequently results in greater movie performance.

Review depth and volume are facilitated and this shows that movies have merits in providing enthusiasm or fun and attracting customers which are not just explainable by large movie resources powers. If people are induced to create online activities, it indicates that the movies have some attractiveness beyond the great profile of actors, directors, distributors, and production companies. Review depth and volume can exert a positive effect on increasing movie efficiency which indicates an efficient use of slack resources involved with box office.

Hypothesis 1: Review depth exerts a positive effect on the production efficiency of the movie.

Hypothesis 2: Review volume exerts a positive effect on the production efficiency of movies.

Review ratings and the number of positive reviews represent the valence or sentiment of reviews and show the evaluation scale offering a persuasion effect to the potential customers. The strength of eWOM valence can have an effect on movie performance because it influences the persuasion effect from the positive feedback mechanism [29,31-33].

With an increase in review rating and a volume containing positive reviews which reflects the attractive value that exists beyond brand powers, movie performance increases in terms of production efficiency. Production efficiency is positively affected by the good evaluation based on review rating and the number of positive reviews.

Hypothesis 3: Review rating exerts a positive effect on the production efficiency of the movie.

Hypothesis 4: The percentage of emotional reviews exerts a positive effect on the production efficiency of movies.

Review helpfulness exerts a positive influence on movie performance by moderating the relationship between eWOM and the box office [30]. Consumers are likely to place greater trust in the contents of reviews if reviews are helpful in that they provide value and consumers are subsequently satisfied with their information needs [33]. Thus, our study posits that review helpfulness exerts a positive effect on the production efficiency of movies. Our study suggests both review and reviewer helpfulness as eWOM factors affecting movie production efficiency. The set of inefficient movies is positioned below the Pareto-efficient frontier, and efficient movies are categorized after comparing movie resource powers with revenue, DEA provides an optimization on each movie and produces the set of Pareto-efficient movies positioning on a piecewise and discrete frontier. When review or reviewer helpfulness increases, there should be a greater likelihood that customers have a positive attitude toward the specific movie [31] and watch the movie, which increases production efficiency and helps inefficient movies move toward the Pareto-efficient frontier.

Hypothesis 5: Review helpfulness exerts a positive effect on the production efficiency of the movie.

Methods

The procedure of this study follows two steps. First, our study collects data regarding eWOM and movie samples from movie websites and performs movie efficiency analysis by evaluating the powers of stars, directors, distributors, and production companies for producing movie revenue

As parts of resources for movies, our study selected four powers, i.e., powers of stars, directors, distributors, and production companies to compute production efficiency as these are closely related to movie revenue. For instance, a featuring of a star positively affects movie revenue [12]. Actors' power is an important determinant of movie revenue [11] indicating that publicity created by celebrities encourages customers to visit cinemas. Further, the power of directors, distributors, and production companies can be considered as affecting determinants for movie revenue, which is assessed using the total revenue created by them.

Previous literature suggests review rating and volume as positive factors for movie revenue [31]. eWOM characteristics also include a percentage of positive or negative eWOM [17,34]. Our study suggests review volume, depth, rating, and sentiment (the number of positive reviews) as determinants for production efficiency.

This study adopts daily ticket sales data which are public online from November 2003 to February 2005. The final sample comprises 1640 movies. The movie-related data and eWOM data were scrawled from the Naver Movies website (see http://movie.naver.com/) between three weeks before release and three weeks after release. The revenues and powers data are compiled from KOFIC (see https://www.kofic.or.kr/) site which is an official site involved with supporting the Korean movie industry. The study selects independent factors adapted from previous studies, which are crucial in many cases. The variables adopted in this study are shown in Table 1.

DEA model is adopted to suggest constant returns and radial improvement for various pairs of inputs (movie resources power) and outputs (movie revenue). The efficiency distributions are suggested in Tables 2 and 3. Table 2 presents that the week 1 after release provides significantly greater efficiency than week 2 and 3. Week 2 shows greater efficiency than week 3. This shows that movie efficiency is decreasing with time due to the decrease in revenue despite the constant movie resources powers.

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015. DOI: https://dx.doi.org/10.17352/amp.S1.000002

Table 3 indicates that in week 1, fantasy movies and Korean movies provide higher efficiency than the other movie classes. There is a variation of efficiency across different classes of movies in utilizing movie resources powers.

This study adopts multivariate regression analysis to test the effect of eWOM on production efficiency. Further, this study uses 12 machine learning approaches k nearest neighbor, k nearest neighbor (bagging), logistic regression, logistic regression (bagging), automated neural networks, neural networks, neural networks (bagging), naïve Bayes, naïve Bayes (bagging), decision trees, decision trees (bagging), random forests.

DEA has been utilized in combination with machine learning methods. For instance, DEA has been adopted in connection with decision support systems for integrating various methods [35], and Tobit analysis for the analysis of efficiency for microfinance institutions [36]. Previous studies have adopted machine learning methods to predict movie revenue using [34]. The statistics-based or machine learning approaches have been used to provide the forecasting box office [9]. Lee, et al. [37] used the Bayesian network to make a forecast of the box office. Our study chooses to utilize DEA with machine learning approaches like k-nearest-neighbors, decision trees, neural networks, and the naïve Bayes method.

Further, this study adopts neural networks, decision trees, k-nearest-neighbors, naïve Bayes method with the methods of combining predictions which build an efficient way to offer better solutions [38,39]. Bagging improves the model's performance stability and facilitates avoiding overfitting by using individual modelings based on separate data samples for combining the results.

In order to make use of neural networks, decision trees, k-nearest-neighbors, naïve Bayes method using bagging, 10 learner models are estimated and integrated to show the box office prediction. This approach builds learning for the preset count of "weak" classification models. The ensemble method is to suggest the combined prediction based on these classification models created.

Experimental results

For the production efficiency at week t, the study uses the eWOM as independent variables at week t-1. Tables 4 and 5 show the multiple regression analysis result using four eWOM variables including review depth, review volume, review rating, and the percent of positive reviews for explaining production efficiency at weeks 1 and 2, respectively. 1671 movies are used in multiple regression analysis.

The moderating effects of review helpfulness can be evaluated using the significance of the interaction terms

Table 1: eWOM and movie-related variables used in this study.

Category	Variables	Description	Number of possible values
eWOM (time-varying)	Review depth (Average length of a review in words)	Indicates the average review length in terms of words	Real values
	Review volume (Average number of reviews)	Indicates the average number of reviews	Real values
	Average review rating	Indicates the average review rating	Real values
	Percent of emotional reviews	Represents the average percent of emotional reviews	Real values
	Helpfulness of review	Shows the average helpfulness of review for each movie	Real values
Movie resources powers	Star power	Indicates the total revenue realized by the actors for each movie	Real values
	Directors power	Indicates the total revenue realized by the directors for each movie	Real values
	Distributors power	Indicates the total revenue realized by the distributors for each movie	Real values
	Production companies power	Indicates the total revenue realized by the production companies for each movie	Real values

Table 2: DEA Efficiency Results.

Box office revenue	Week 1	Week 2	Week 3	Week 1 – 2		Week 1 – 3		Week 2 – 3	
				t	Sig.	t	Sig.	t	Sig.
Number of efficient movies	475	416	393	9.88	0.000	10.65	0.000	4.31	0.000
Number of inefficient movies	1195	1255	1278						
Average efficiency (%)	44.30	36.84	33.03						

Table 3: Distributions of efficiency.

Category	Value	# of movies	Total revenue (X 10^9 Won)	Average efficiency	
				Week 1	Week 2
Genre	History	3	16.0	12.7	6.5
	Drama	775	1.79	38.9	33.1
	Thriller/mystery	59	2.33	50.1	33.5
	Horror	67	1.47	54.1	45.3
	Mello/Romance	180	1.56	44.6	36.2
	Animation	104	1.86	56.0	46.9
	Comedy	207	2.24	44.9	37.7
	War	1	0.007	0,85	1.99
	Action	118	4.96	48.8	37.0
	Science fiction	11	7.91	50.0	34.0
	Documentary	107	0.084	55.0	51.9
	Art	4	0.79	51.0	40.8
	Crime	23	3.47	47.1	39.3
	Fantasy	4	0.77	42.3	10.2
	Family	7	1.10	55.4	26.3
Nation	Korea	663	3.26	51.6	46.0
	US	101	0.45	39.6	31.4
	England	77	0.52	36.6	22.9
	France	458	2.15	42.8	31.7
	Japan	153	0.43	34.0	29.6

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015. DOI: https://dx.doi.org/10.17352/amp.S1.000002

Table 4: Regression analysis on movie efficiency of the first week after release F = 7.05, *p* = 0.000.

Independent variables	Standardized Coefficients	t	Sig.
Review depth (prerelease)	-.060	-.697	.486
Review volume (prerelease)	.319	3.296	.001
Review rating (prerelease)	-.141	-1.719	.086
The percent of emotional reviews (prerelease)	.083	2.202	.028
Review depth (prerelease) X helpfulness (prerelease)	.065	.412	.680
Review volume (prerelease) X helpfulness (prerelease)	-.183	-1.893	.059
Review rating (prerelease) X helpfulness (prerelease)	.351	2.057	.040
The percent of emotional reviews (prerelease) X helpfulness (prerelease)	.015	.120	.905

Table 5: Regression analysis on movie efficiency of the second week after release F = 4.34, *p* = 0.000.

Independent variables	Standardized Coefficients	t	Sig.
Review depth (first week)	.008	.098	.922
Review volume (first week)	.247	1.705	.089
Review rating (first week)	-.034	-.470	.638
The percent of emotional reviews (first week)	-.017	-.371	.711
Review depth (first week) X helpfulness (first week)	-.174	-1.081	.280
Review volume (first week) X helpfulness (first week)	-.083	-.570	.569
Review rating (first week) X helpfulness (first week)	.409	1.769	.077
The percent of emotional reviews (first week) X helpfulness (first week)	.169	1.585	.113

between eWOM and review helpfulness in multivariate regression analyses where the dependent variable is movie revenue. Tables 4 and 5 present regression analysis on the production efficiency of the first and second week after release where the interaction term is about one week before.

Tables 4 and 5 show that helpfulness provides a positive moderating effect with review rating on efficiency for weeks 1 and 2. Review volume positively affects production efficiency while review rating has an interaction effect on production efficiency moderated by helpfulness. This indicates that review volume is more directly influencing production efficiency than review rating or depth. The indirect effect of review rating through moderating by helpfulness is greater than for review depth or volume. These results of direct and moderating effects of eWOM on production efficiency show that the review volume is an important factor for determining production efficiency and that review rating can have an influence on production efficiency when there exists a high level of review helpfulness.

The results of the study present that the slack resources of inefficient movies are able to be negatively associated with the number of reviews and the percentage of emotional reviews Review rating is influenced by helpfulness to affect production efficiency or the efficient use of movie resources powers than the other three eWOM factors depth, volume, and percent of emotional reviews. The greater efficiency represents less slack resources and the review rating is less related to the slack resources of movie production. Because production efficiency shows that a movie is successful even when resources powers are not much used, other factors exist that can more than compensate the brand powers of actors, directions, distributors, and production companies. These encompass peculiar or unique merits in movie technology, scenario, and composition which can attract customers. This results in the moderating effects of review helpfulness for the effect of review rating on production efficiency. In the moderating relation, review helpfulness exerts a role of credibility and the greater the review helpfulness, the more effectively and correctly showing the real success of movies in terms of the relationship between movie budget and box office.

On the contrary, review volume has a negative moderating effect on review helpfulness, which shows that the number of reviews can be a misleading measure for predicting production efficiency while it has a positive direct effect on efficiency. Consequently, the moderating effects on movie efficiency by eWOM variables are negative for volume and positive for review rating.

The negative moderating effect of volume interacted by review helpfulness on production efficiency shows that the slack resources of inefficient movies can be positively influenced by review volume when it is moderated by review helpfulness. Production efficiency represents that the box office can be resulted even when there exist not so much spent resources, other merits exist to bring out revenue like scenario composition, and special movie effects to create great response from movie customers. Thus, review rating and the percentage of emotional reviews are more influencing efficiency through interaction effects with review helpfulness. The moderation effects on production efficiency by review rating or review sentiment moderated by review helpfulness can be positively significant.

Our study uses training data to perform learning of parameters for the neural networks, decision trees, k-nearest-neighbors, and naïve Bayes method to forecast the value of production efficiency in each target record in a validation sample. This study provides the prediction function learned from the training data and adopts non-statistical and statistical models to perform the box-office prediction. This study applies four business intelligence methods (neural networks, decision trees, k-nearest-neighbors, naïve bayes) using bagging methods. The seven independent variables are used: review depth, volume, rating, three sentiments (average number of emotional reviews, average number of positive reviews, average number of negative reviews), average review extremity (absolute difference between review rating and average rating).

The sample of 1671 records is divided into a training sample and a validation sample. The sample was divided into 34 subsets each of which has 50 (the last one has 21 records) movies, respectively. Each of these subsets is used as a validation sample one by one, the movies left over are adopted as training samples. That is, for assessing the comparison of the performance of prediction at week 1, training and validation

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015. DOI: https://dx.doi.org/10.17352/amp.S1.000002

samples are built in 34 pairs each having 1621 and 50 movies. The n-fold cross-validation of samples and three dependent variables of production efficiency (at each of three weeks) is adopted to provide the stability of the prediction results for the comparison of 12 data mining methods toward the prediction for movie efficiency in the corresponding validation sample. The prediction error is produced using classification errors. After the classified value is compared with the true value in the validation sample to produce a classification error, these prediction errors are then averaged across 34 validation samples.

Table 6 shows the average classification errors in ascending order when eWOM factors are used as independent variables to classify production efficiency across weeks 1, 2, and 3. Table 6 shows that the average classification errors are lowest when k nearest neighbor, logistic regression, logistic regression with bagging, and automated neural networks are used as the classification method across weeks 1, 2, and 3. Tables 7 and 8 indicate t-test results for the difference in the average classification errors according to k nearest neighbor and automated neural networks, respectively. Tables 7 and 8 indicate that the average classification errors are significantly lower when k nearest neighbor and automated neural networks are used as the classification method. These results show that k nearest neighbor and automated neural networks outperform the other machine learning methods in classifying efficient movies.

Discussion

This study presents that besides box office, production efficiency can be used as a potential measure of movie performance. The marketing function can use efficiency measures to assess the performance of a movie and can produce a prediction of this using eWOM factors as independent factors. Marketing managers could utilize movie efficiency with respect to the box office in terms of how much they accomplish box office by minimizing the usage of movie resources and figuring out the effect of eWOM such as review volume, rating, and sentiment on efficiency. These eWOM factors can be evaluated one week before the movie release and can be used to make a forecasting of efficiency one week later.

Our study builds insights for the current movie literature on eWOM by suggesting the effect of interaction between eWOM

Table 6: Average classification errors across 34 test samples based on various machine learning methods.

Methods \ Week	First week	Second week	Third week
k nearest neighbor	0.2765	0.2465	0.2361
Logistic regression	0.2782	0.2432	0.2302
Logistic regression (bagging)	0.2794	0.2438	0.2302
Automated neural networks	0.2832	0.2479	0.2320
Neural networks	0.2832	0.2479	0.2320
Neural networks (bagging)	0.2832	0.2491	0.2320
Naïve bayes (bagging)	0.2882	0.2644	0.2473
Naïve bayes	0.2915	0.2626	0.2414
Decision trees (bagging)	0.3026	0.2562	0.2461
Decision trees	0.3071	0.2694	0.2659
Random forests	0.3100	0.2544	0.2391
k nearest neighbor (bagging)	0.3803	0.3788	0.3410

Table 7: t-test of the difference in average classification error (with respect to k nearest neighbor).

Methods	Difference in average classification error	*t* value	Significance
k nearest neighbor for wk1 – k nearest neighbor (bagging) for wk1	-.1038	-7.355	.000
k nearest neighbor for wk1 – Random forests for wk1	-.0335	-4.240	.000
k nearest neighbor for wk1 – Decision trees for wk1	-.0306	-3.636	.001
k nearest neighbor for wk1 – Decision trees (bagging) for wk1	-.0262	-2.807	.008
k nearest neighbor for wk1 – Naïve Bayes for wk1	-.0150	-1.862	.072
k nearest neighbor for wk2 – k nearest neighbor (bagging) for wk2	-.1324	-10.894	.000
k nearest neighbor for wk2 – Decision trees for wk2	-.0229	-2.846	.008
k nearest neighbor for wk2 – Naïve Bayes for wk2	-.0162	-1.911	.065
k nearest neighbor for wk2 – Naïve bayes (bagging) for wk2	-.0179	-2.079	.045
k nearest neighbor for wk3 – k nearest neighbor (bagging) for wk3	-.1049	-10.966	.000
k nearest neighbor for wk3 – Decision trees for wk3	-.0298	-3.393	.002

Table 8: t-test of the difference in average classification error (with respect to automated neural networks).

Methods	Difference in average classification error	*t* value	Significance
Automated neural networks for wk1 – k nearest neighbor (bagging) for wk1	-.0971	-5.999	.000
Automated neural networks for wk1 – Random forests for wk1	-.0268	-4.229	.000
Automated neural networks for wk1 – Decision trees for wk1	-.0238	-3.554	.001
Automated neural networks for wk1 – Decision trees (bagging) for wk1	-.0194	-3.302	.002
Automated neural networks for wk2 - k nearest neighbor (bagging) for w2	-.1309	-7.759	.000
Automated neural networks for wk2 – Decision trees for wk2	-.0215	-2.650	.012
Automated neural networks for wk2 – Naïve Bayes for wk2	-.0147	-2.342	.025
Automated neural networks for wk2 – Naïve bayes (bagging) for wk2	-.0165	-2.746	.010
Automated neural networks for wk3 – k nearest neighbor (bagging) for wk3	-.1090	-9.528	.000
Automated neural networks for wk3 – Random forests for wk3	-.0071	-1.977	.056
Automated neural networks for wk3 – Decision trees for wk3	-.0339	-3.987	.000
Automated neural networks for wk3 – Decision trees (bagging) for wk3	-.0141	-2.564	.015
Automated neural networks for wk3 – Naïve Bayes for wk3	-.0094	-2.610	.014
Automated neural networks for wk3 – Naïve bayes (bagging) for wk3	-.0153	-2.862	.007

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015. DOI: https://dx.doi.org/10.17352/amp.S1.000002

(review rating) and helpfulness on production efficiency. This is to extend previous literature regarding the effect of eWOM on box office or product sales [40,41]. Production efficiency indicates the efficient usage of movie resources, and review rating moderated by review helpfulness affects movie performance in terms of efficiency. In addition, review volume and sentiment can indicate that movie performance is production efficiency. Our study presents that different eWOM variables (volume *vs.* rating) can be differently (negatively or positively) moderated in direction by helpfulness to affect production efficiency.

This study further shows that appropriate business analytics approaches can be selected based on classification errors. Or study shows that k nearest neighbor, logistic regression, and automated neural networks are the best methods to make a forecasting for production efficiency. This provides insights into box office prediction studies because it shows the guidance in selecting the best data mining approaches for predicting movie efficiency.

Conclusion

The study offers several implications to research on the box office in that it suggests explanatory analytics of eWOM factors on production efficiency which is not well studied in previous studies on movies (e.g., [15]) and intends to make the prediction of efficiency using eWOM factors. Movie resource powers are critical factors for the box office [10-12], and our study employs movie efficiency in using resource powers for the box office and offers insights into movie research by examining how eWOM factors affect movie efficiency and whether there exist moderating effects on production efficiency between eWOM factors and review helpfulness. Our study extends the former research on resource powers by using movie efficiency by comparing four resource powers, i.e., powers of actors, directors, distributors, and production companies, as a whole with box office based on the DEA approach to show how much resources are invested in movie production and how movie efficiency is related to eWOM factors like review volume rating, and sentiment.

Our study employs predictive analytics methods to examine their predictive performance in predicting production efficiency. It is insightful to test the predictive efficacy of data mining methods for this alternative measure of movie performance as the studies on the prediction of movie production efficiency are almost nonexistent. The results show that k nearest neighbor and automated neural networks outperform the other machine learning methods in classifying efficient movies.

Our study has several limitations and future research issues. First, our study utilizes production efficiency as the promising alternative measure of movie performance, and there can exist other causal independent factors or movie-related variables to be moderators to affect production efficiency which is a relatively new concept in movie literature. Second, the moderating role of review helpfulness needs to be further investigated in the other context of the data set. Third, the efficiency measure can be further developed based on other resources powers, or movie budget. Future studies can further investigate the relationship between eWOM and new measures of production efficiency.

Funding

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A2A03052219).

References

1. MyPlan, Industry Outlook, Motion Picture & Video Industries. https://www.myplan.com/careers/industries/motion-picture-and-video-industries/industry -outlook-nature-of-the-industry-06.04.html. 2020.
2. Liu T, Ding X, Chen Y, Chen H, Guo M. Predicting movie Box-office revenues by exploiting large-scale social media content. Multimedia Tools and Applications. 2014; 75:509–1528.
3. Eliashberg J, Elberse A, Leenders, MA. The motion picture industry: Critical issues in practice, current research, and new research directions. Marketing science. 2006; 25(6):638-661.
4. Lee S, Choeh JY. Movie production efficiency moderating between online word-of-mouth and subsequent box office revenue. Sustainability. 2020; 12(16):6602.
5. Lee S, Kc B, Choeh JY. Comparing performance of ensemble methods in predicting movie box office revenue. Heliyon. 2020 Jun 20;6(6):e04260. doi: 10.1016/j.heliyon.2020.e04260. PMID: 32613125; PMCID: PMC7322254.
6. Madongo CT, Zhongjun T. A movie box office revenue prediction model based on deep multimodal features. Multimedia Tools and Applications. 2023; 82:31981–32009.
7. Velingkar G, Varadarajan R, Lanka S, A.K.M. Movie box-office success prediction using machine learning. 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India. 2022; 1-6.
8. Wang Z, Zhang J, Ji S, Meng C, Li T, Zheng Y. Predicting and ranking box office revenue of movies based on big data. Information Fusion. 2020; 60:25-40.
9. Zhang L, Luo J, Yang S. Forecasting box office revenue of movies with BP neural network. Expert Systems with Applications. 2009; 36(3): 6580–6587.
10. Pangarker NA, Smit EvdM. The determinants of box office performance in the film industry revisited. South African Journal of Business Management. 2013; 44(3):47-58.
11. Treme J. Effects of celebrity media exposure on box-office performance. Journal of Media Economics. 2010; 23:5-16.
12. Treme J. Craig, L.A. Celebrity star power: do age and gender effects influence box office performance? Applied Economics Letters. 2013; 20:440–445.
13. Asadpour M, Shirouyehzad H. Performance evaluation and ranking of Academy Award winners for best original score applying data envelopment analysis: 1990-2016. Operations Research Letters. 2019; 47(5):371-376.
14. Hababou M, Amrouche N, Jedidi K. Measuring economic efficiency in the motion picture industry: a data envelopment analysis approach. Customer Needs and Solutions. 2016; 3(3-4):144-158.
15. Lee S, Choeh JY. The moderating effect of movie production efficiency on the relationship between eWOM and review helpfulness. Management Research Review. 2021; 43(11):341-358.
16. Du J, Xu H. Huang, X. Box office prediction based on microblog. Expert Systems with Applications. 2014; 41:1680–1689.

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015. DOI: https://dx.doi.org/10.17352/amp.S1.000002

17. Rui H, Liu Y, Whinston A. Whose and what chatter matters? The effect of tweets on movie sales. Decision Support Systems. 2013; 55:863–870.

18. Hennig-Thurau T, Wiertz C, Feldhaus F. Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. Journal of the Academy of Marketing Science. 2015; 43(3): 375-394.

19. Marchand A, Hennig-Thurau T, Wiertz C. Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. International Journal of Research in Marketing. 2017; 34(2):336-354.

20. Rosario AB, Sotgiu F, De Valck K, Bijmolt TH. The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. Journal of Marketing Research. 2016; 53(3):297-318.

21. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research. 1978; 2(6):429–444.

22. Emrouznejad A, Anouze, AL. Data envelopment analysis with classification and regression tree –A case of banking efficiency. Expert Systems. 2010; 27(4):231–246.

23. Cooper WW, Seiford LM, Zhu J. Data envelopment analysis: History, models and interpretations. In WW Cooper, LM. Seiford, J Zhu (Eds.), Hand- book on data envelopment analysis. Dordrecht: Kluwer Academic Publishers Group. 2004; 1–39.

24. Lee S. Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls. Decision Support Systems. 2010; 49(4):486-497.

25. Huerta TR, Thompson MA, Ford EW, Ford WF. Electronic health record implementation and hospitals' total factor productivity. Decision Support Systems. 2013; 55(2):450-458.

26. Vandaele NJ, Decouttere CJ. Sustainable R&D portfolio assessment. Decision Support Systems. 2013; 54(4):1521-1532.

27. Chou YC, Shao BBM. Total factor productivity growth in information technology services industries: A multi-theoretical perspective. Decision Support Systems. 2014; 62:106-118.

28. Chevalier JA, Mayzlin D. The effect of word of mouth on sales: online book reviews. Journal of Marketing Research. 2006; 43(3):345-54.

29. Liu Y, Word of mouth for movies: Its dynamics and impact on box office revenue. Journal of Marketing. 2006; 70:74-89.

30. Lee S, Choeh JY. The impact of online review helpfulness and word of mouth communication on box office performance predictions. Humanities & Social Sciences Communications. 2020; 7:84.

31. Chintagunta PK, Gopinath S, Venkataraman S. The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. Marketing Science. 2010; 29(5):944–957.

32. Wu J. Review popularity and review helpfulness: a model for user review effectiveness. Decision Support Systems. 2017; 97: 92-103.

33. Filieri R, Alguezaui S, McLeay F. Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism Management. 2015; 51:174–185.

34. Kim T, Hong J, Kang P. Box office forecasting using machine learning algorithms based on SNS data. International Journal of Forecasting. 2015; 31:364–390.

35. Samoilenko S, Osei-Bryson KM. A hybrid DEA-/DM-based DSS for pro- ductivity-driven environments. In K. M. Osei-Bryson, & O. Ngwenyama (Eds.), Advances in research methods for information systems research. New York: Springer Science + Business Media. 2014; 165–181.

36. Widiarto I, Emrouznejad A, Anastasakis L. Observing choice of loan methods in not-for-profit microfinance using data envelopment analysis. Expert Systems with Applications. 2017; 82(1):278-290.

37. Lee S, Lee KC, Choeh JY. Using Bayesian network to predict online review helpfulness. Sustainability. 2020; 12(17):6997.

38. Shmueli G, Bruce P, Gedeck P, Patel N. Data mining for Business Analytics: Concepts, Techniques, and Applications in Python, John Wiley & Sons, Inc. 2020.

39. Lee S, Choeh JY. The interactive impact of online word-of-mouth and review helpfulness on box office revenue. Management Decision. 2018; 56(4):849-866.

40. Zhu F, Zhang X. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing. 2010; 74:133–148.

Citation: Lee S, Choeh JY (2023) Explanatory and predictive analytics for movie production efficiency by online word-of-mouth. Ann Math Phys S1(1): 008-015. DOI: https://dx.doi.org/10.17352/amp.S1.000002