

Modeling Influencing Factors in U.S. Film Success (1940-2024)

Jinglin Xiao

College of Arts & Sciences, University of Washington, Seattle, USA

Email: jx77@uw.edu

How to cite this paper: Xiao, J. L. (2024). Modeling Influencing Factors in U.S. Film Success (1940-2024). *Modern Economy*, 15, 1319-1334.

<https://doi.org/10.4236/me.2024.1512068>

Received: October 20, 2024

Accepted: December 20, 2024

Published: December 23, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study explores the economic impact of marketing and casting strategies on the U.S. film industry, valued at almost \$93 billion in 2022. With high production costs and shifting consumer preferences, effective marketing and casting choices are crucial for maximizing return on investment (ROI). Using different models: Logistic regression, Random Forest and SVM models in R, this research analyzes factors like release timing, critic ratings, and genres to predict box office success, defined by high revenue and awards. The Random Forest Model achieved the highest accuracy, emphasizing that variables like “Runtime”, “IMDb Rating”, and “Rotten Tomatoes Rating” are critical predictors of success. The findings highlight the importance of strategic marketing that optimizes resources without exceeding diminishing returns. Furthermore, the study acknowledges data limitations caused by missing film details, such as those for *Avatar: The Way of Water*, which affect the completeness of the dataset. Future research should incorporate streaming data to better assess film success beyond box office metrics. These insights offer industry stakeholders data-driven recommendations for optimizing marketing, casting, and release strategies in an evolving entertainment landscape.

Keywords

Film Industry, Marketing Strategies, Casting Choices, Box Office Success, Predictive Modeling

1. Introduction

The US film industry, a significant contributor to the global economy, was valued at “almost \$93 billion” as of 2022 (Carollo, 2024). Beyond its cultural influence, the industry plays a crucial role in shaping economic trends through employment generation, consumer spending, and international trade. One of the most critical

economic factors influencing a movie's success is its strategic marketing and casting decisions. These decisions can substantially impact a film's financial performance, determining its ability to recoup production costs and make profit. As this study focuses on the profitability, a movie's success is defined by high box office performance and/or high recognition through awards. This paper seeks to explore the economic implications of marketing strategies and casting choices, examining how they can affect a film's market demand, revenue potential, and overall success in the competitive entertainment landscape.

The core problem addressed in this study is the economic effectiveness of marketing and casting strategies in the film industry. In a market characterized by high production costs and varying consumer preferences, filmmakers and studios face the challenge of maximizing return on investment (ROI) through effective resource allocation. Strategic marketing campaigns, such as targeted advertising, promotional partnerships, and social media engagement, can significantly influence consumer demand and willingness to pay (Skye, 2024). Similarly, casting decisions, particularly involving bankable stars or directors, can enhance a film's marketability, directly impacting its revenue-generating potential. The question lies in understanding which factors yield the highest economic returns and how these can be optimized for different market conditions.

The objectives of this paper are to:

- 1) Identify and evaluate the economic impact of different marketing strategies and casting choices on a film's success.
- 2) Analyze the relationship between key attributes (release timing, critic and audience ratings, genres, and production budgets) and box office performance.
- 3) Provide data-driven recommendations to aid industry stakeholders in making informed decisions regarding film production, marketing, and release strategies.

The study's scope focuses on U.S. movies released after 1940, offering a historical and contemporary view of how marketing strategies and casting choices have influenced film success over time. This broad time frame allows for a comprehensive economic analysis across various eras and genres, including the evolution of audience preferences and market trends.

The study employs various models—Logistic Regression, Random Forest, and Support Vector Machine (SVM)—implemented in RStudio to conduct an in-depth analysis of the primary factors influencing U.S. movie box office performance. These factors include release timing (e.g., holiday seasons), Rotten Tomatoes scores, audience ratings, film genres, production budgets, revenue, and other relevant variables. By systematically analyzing these predictors, the study aims to assess their impact on box office performance and industry recognition. This approach examines the relationships between these key attributes and film success, providing data-driven recommendations to assist industry stakeholders in making more informed decisions regarding film production, marketing, and release strategies.

2. Literature Review

The U.S. film industry significantly impacts the national economy through box office revenues, streaming services, and merchandising (Nichols, 2018). In 2023, U.S. box office revenues totaled approximately \$9 billion, underscoring the industry's economic relevance (Domestic Box Office, 2023). Effective marketing and casting are pivotal for a film's profitability, with studies examining their influence on financial performance.

2.1. Theoretical Framework

Cultural economics frames films as unique cultural goods shaped by consumer preferences, distribution strategies, and cultural value. **Star power theory**, emphasized by Lash and Zhao (2016), highlights the role of well-known actors in attracting audiences and boosting box office performance. Chisholm et al. (2014) also underscore casting's economic significance.

2.2. Marketing Strategies in the Film Industry

Traditional marketing channels like TV and print create broad exposure (Scott, 2019). Digital marketing has emerged as a key strategy, leveraging social media platforms to target audiences effectively (He & Hu, 2021). While large budgets typically correlate with higher revenues, diminishing returns occur when market saturation is reached (Wisnefsky, 2023; Lash & Zhao, 2016).

2.3. Casting Choices and Star Power

Star power, measured by actors' previous earnings and media presence, drives strong opening weekends (Lash & Zhao, 2016). However, casting decisions must balance popularity with narrative quality to sustain success (He & Hu, 2021). Excessive reliance on star power can also result in diminishing returns (Wisnefsky, 2023).

2.4. Methodological Approaches in Film Industry Research

2.4.1. Statistical Models and Machine Learning

Research on film success prediction frequently employs statistical and machine learning methods. Scott (2019) used regression analysis to examine factors like marketing spend, star power, and release timing on box office revenue. He and Hu (2021) utilized machine learning techniques to analyze large datasets, offering improved predictive accuracy. Social network analysis by Lash and Zhao (2016) highlighted relationships between actors, directors, and audiences as key profitability indicators.

2.4.2. Challenges in Data Collection

Data inconsistencies for international films and independent productions present challenges (Chisholm et al., 2014). Additionally, subjective metrics such as reviews and audience sentiment are difficult to quantify, complicating their integration into models (Lash & Zhao, 2016). These issues can limit the generalizability of

findings.

2.4.3. Factors Influencing Box Office Performance

- **Intrinsic Factors:** Characteristics like genre, MPAA rating, and sequels significantly influence performance. Family-friendly films (G/PG-rated) outperform others, while sequels benefit from established fanbases (Scott, 2019). Action and superhero genres also attract larger audiences (Lash & Zhao, 2016).
- **Release Timing and Critical Reception:** Timing is crucial, with summer and holiday releases typically outperforming those in less competitive periods (Scott, 2019). Wider releases on more screens also boost revenue (He & Hu, 2021). Positive reviews on platforms like Rotten Tomatoes and strong word-of-mouth drive sustained success (He & Hu, 2021).

2.5. International Markets and Emerging Trends

Films must adapt marketing strategies to cultural contexts. For instance, Chinese audiences prefer action-packed spectacles, while European audiences favor narrative-driven films (Lash & Zhao, 2016). Additionally, streaming platforms like Netflix challenge traditional box office metrics while introducing new revenue opportunities (He & Hu, 2021).

2.6. Gaps and Future Research Directions

2.6.1. Identifying Gaps

While existing research examines marketing spend, casting, and timing, gaps remain regarding the impact of streaming platforms and digital metrics on film success. Further studies should explore diminishing returns on marketing in the streaming era and the interplay between cultural factors, traditional, and digital marketing strategies.

2.6.2. Positioning This Study

The study addresses these gaps by analyzing traditional and digital marketing strategies, casting decisions, and other film attributes. By integrating prior findings, it provides actionable insights and a comprehensive perspective on film success, contributing to the evolving understanding of the industry's economic dynamics.

3. Methodology

3.1. Data Collection and Sources

This study examines factors influencing U.S. movie success, focusing on release timing (e.g., holiday seasons) and audience engagement, measured by the vote-to-popularity ratio, to analyze their correlation with return on investment (ROI).

Data was collected from two main sources: The Movie Database (TMDb), offering details like runtime, budget, revenue, and popularity, and the Open Movie Database (OMDb), providing additional metrics, including Rotten Tomatoes and IMDb ratings. Together, these sources created a comprehensive dataset integrat-

ing qualitative and quantitative film performance indicators.

3.2. Data Cleansing and Feature Engineering

Prior to analysis, as shown in **Figure 1**, the dataset underwent thorough cleansing and feature engineering to ensure data integrity and enhance model performance:

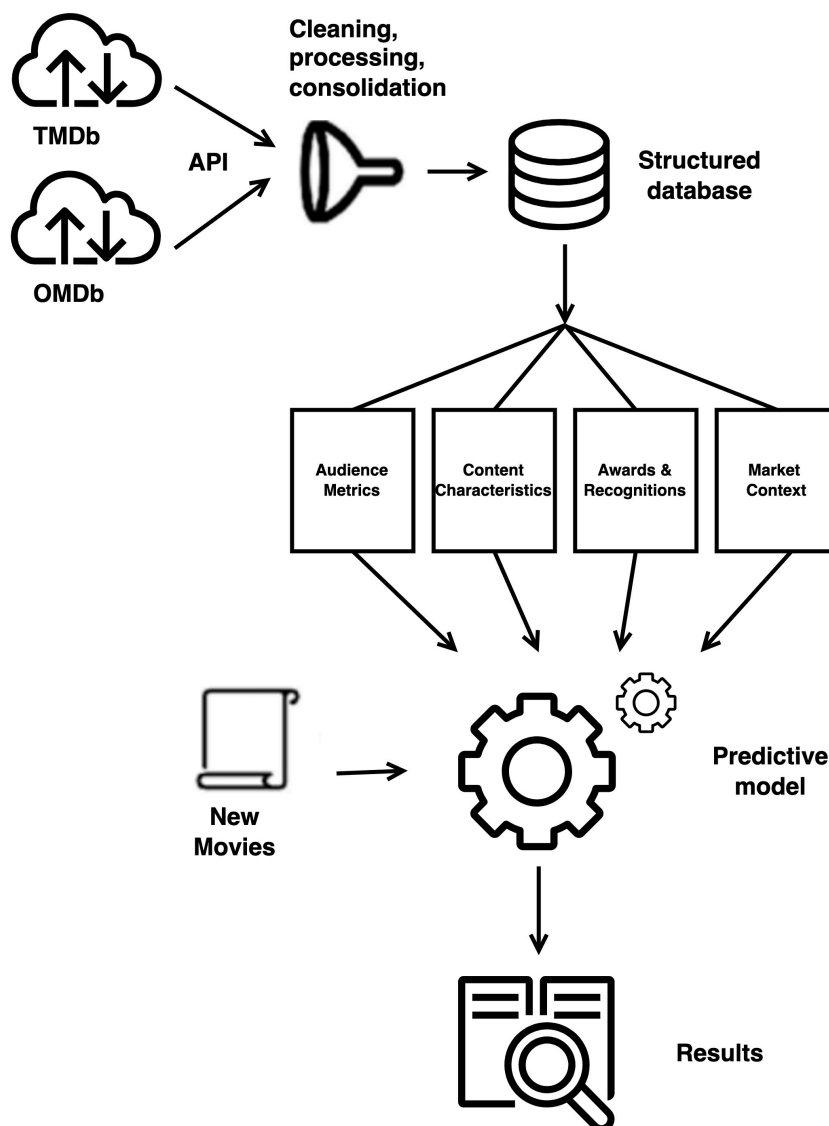


Figure 1. The framework of the study.

1) **Missing Data:** Key fields like budgets and revenues were imputed or removed.

2) **Feature Selection:** Relevant features such as IMDb ratings and holiday releases were retained to reduce multicollinearity.

3) **Balancing:** SMOTE addressed imbalances between successful and non-successful films.

4) **Splitting:** Data was split 70:30 into training and testing sets (**Figure 2**).

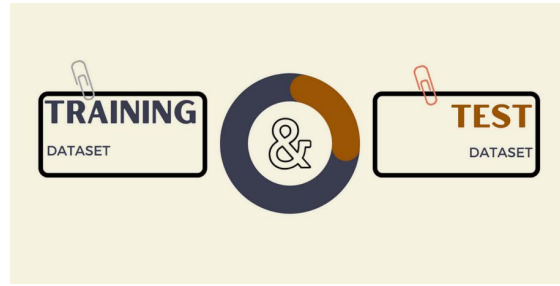


Figure 2. The train-test split technique.

3.3. Defining Success

Initially, a movie was classified as profitable if its revenue exceeded its production budget. However, considering that additional costs such as marketing and distribution typically add 25% - 50% to the production budget, the ROI threshold was raised to 1.5. This adjustment ensures that an ROI around 1 likely only covers costs, allowing for a more realistic assessment of a film’s profitability (see Figure 3). Star power was measured by cumulative box office earnings, with actors surpassing \$900 million classified as superstars.

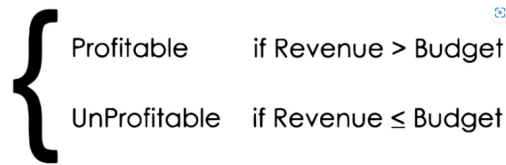


Figure 3. Defining profitability (Mitchellmarino, n.d.).

3.4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to analyze the dataset’s distribution, patterns, and outliers. Python and R were used for data manipulation and visualization, which informed handling missing data and refining the dataset.

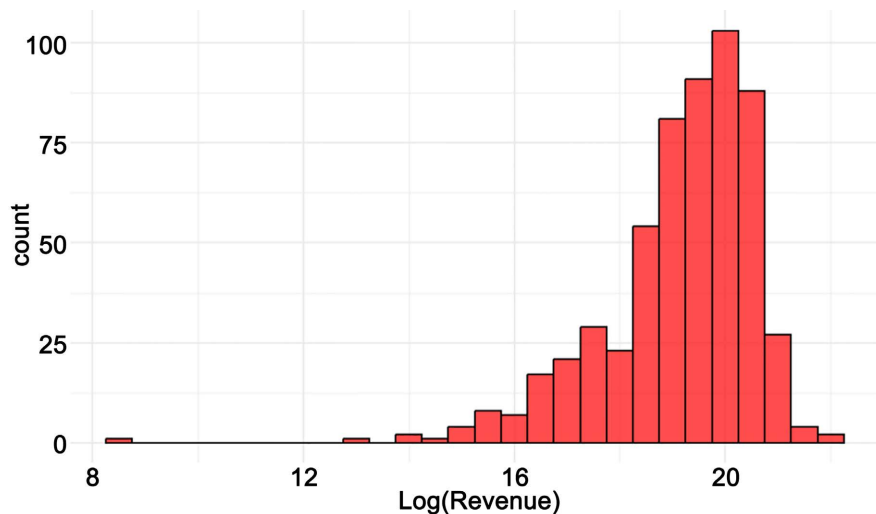


Figure 4. Log-transformed revenue distribution.

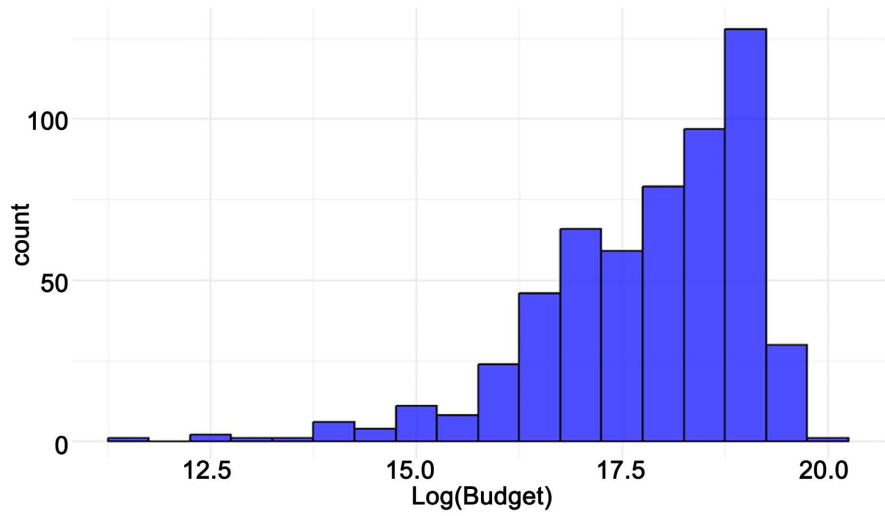


Figure 5. Log-transformed budget distribution.

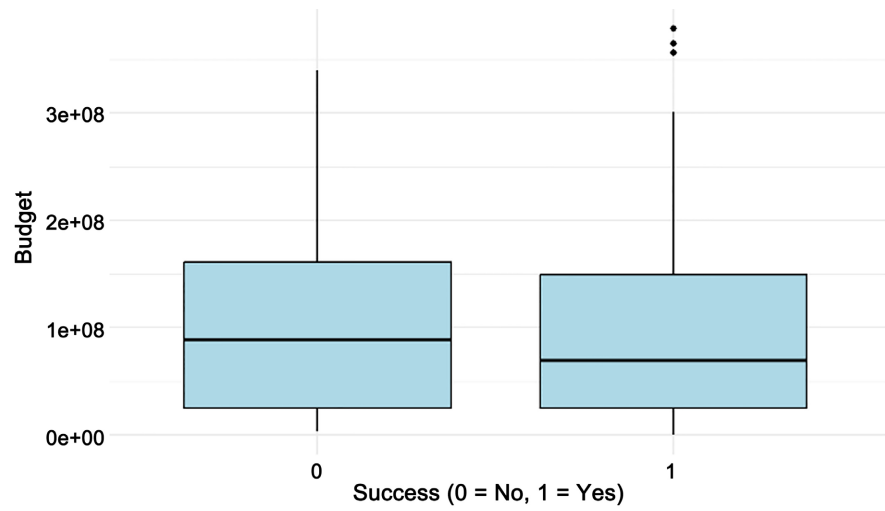


Figure 6. Boxplot of budget by success.

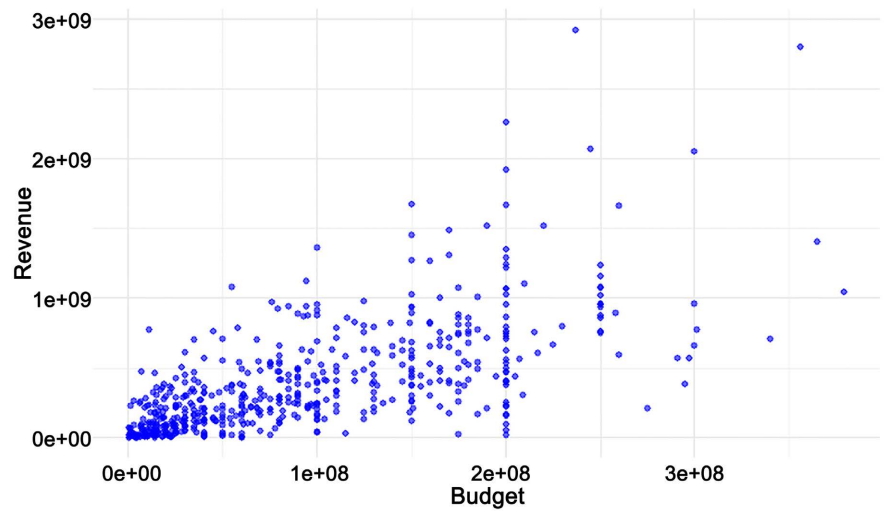


Figure 7. Scatter plot of budget vs revenue.

Log-transformed revenues (Figure 4) and budgets (Figure 5) revealed clearer patterns, with most revenues between log values of 15 - 21 and budgets between 17.5 - 20. Figure 6 compares budgets of successful and non-successful films, showing overlap that suggests budget alone does not predict success. Figure 7 highlights the correlation between budgets and revenue, indicating higher budgets generally yield more revenue but are not guaranteed. The correlation matrix (Figure 8) shows moderate positive correlations between IMDb/Rotten Tomatoes Ratings and success, while “Is a Holiday” and “Has Won Award” have weaker correlations.

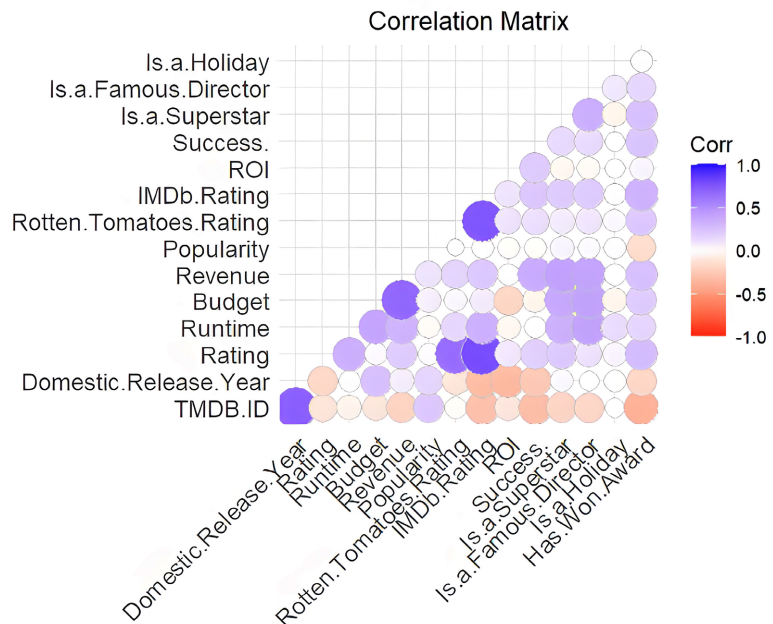


Figure 8. Correlation matrix of movie features and success factors.

3.5. Handling of Multicollinearity

To enhance model reliability, multicollinearity was addressed using:

- VIF values were calculated for each predictor variable using the vif() function from the **car** package in R. A VIF threshold of 5 was established to identify variables exhibiting high multicollinearity, following guidelines by O’Brien (2007). Variables with VIF > 5 were removed.

Table 1. Variance inflation factor (VIF) values for predictor variables.

Predictor	VIF
Has.Won.Award	1.099
Is.a.Superstar	1.166
IMDb.Rating	3.119
Rotten.Tomatoes.Rating	2.786
Runtime	1.324

The VIF values for the final set of predictor variables are presented in **Table 1**. All variables have VIF values below the threshold, indicating acceptable levels of multicollinearity.

- A Pearson correlation matrix was generated to assess pairwise correlations between predictor variables. Correlation coefficients exceeding **0.7** were flagged as indicative of strong linear relationships, following guidelines by **Hair et al. (2010)**.

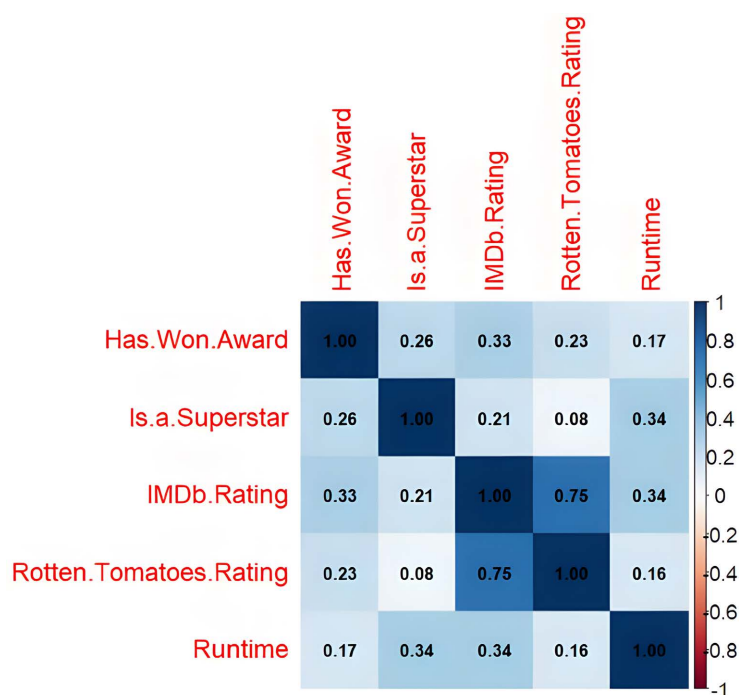


Figure 9. Correlation matrix of movie features.

The correlation matrix (**Figure 9**) indicates that the highest correlation between any two predictors is 0.751 (between IMDb.Rating and Rotten.Tomatoes.Rating). While this exceeds the common threshold of 0.7, it is essential to consider the context and the specific roles these variables play in the model. Both ratings provide distinct insights into film success—IMDb.Rating reflects audience ratings, whereas Rotten.Tomatoes.Rating aggregates critical reviews. Given their unique contributions and the overall low multicollinearity indicated by VIF values, both variables were retained in the final feature set to capture diverse aspects of film success.

As a result, the final feature set included: Has.Won.Award, Is.a.Superstar, IMDb Rating, Rotten Tomatoes Rating, and Runtime.

3.6. Modeling Approach

Three machine learning models were employed to predict film success as a binary outcome (0 = No, 1 = Yes):

- 1) **Logistic Regression (LR):** A fundamental statistical method valued for its

simplicity and interpretability, LR is ideal for understanding relationships between predictor variables and binary outcomes. It serves as a robust baseline for comparing the performance of more complex models.

2) Random Forest (RF): An ensemble method combining multiple decision trees to handle non-linearity, assess feature importance, and reduce overfitting, making it well-suited for datasets with diverse features.

3) Support Vector Machine (SVM): Effective in high-dimensional spaces, SVM is robust against outliers and excels in classification tasks, making it beneficial for distinguishing between successful and non-successful films.

Cross-Validation ensured robust performance by dividing the dataset into training and testing subsets, enabling reliable model comparisons and minimizing overfitting risks.

3.6.1. Rationale for Model Selection

Logistic Regression: Provides clear interpretability and serves as a baseline model to benchmark against more complex approaches.

Random Forest: Handles non-linear relationships effectively, includes mechanisms for evaluating feature importance, and is robust against overfitting due to ensemble learning.

Support Vector Machine: Excels in handling high-dimensional datasets, demonstrates robustness to outliers, and performs strongly in binary classification tasks.

3.6.2. Consideration of Alternative Models

Decision Trees: While intuitive and easy to implement, single decision trees are prone to overfitting and lack the predictive power of ensemble approaches like Random Forest. Their performance limitations make them less suitable for this study's objective of robustly predicting film success.

4. Result

The study's key findings demonstrate the effectiveness of different machine learning models in predicting movie success. The **Random Forest Model** achieved the highest performance, with an accuracy of 81.85% and a sensitivity of 79.41% for class 0. This ensemble model's robustness stems from its ability to combine multiple decision trees, perform random sampling (bootstrapping), and select features to reduce correlation, effectively aggregating outputs for reliable predictions (see **Figure 10**).

The **Support Vector Machine (SVM)** Model demonstrated a balanced accuracy of 74.52% with a sensitivity of 74.26%. While effective for classification tasks, its performance was slightly below the Random Forest Model. Nonetheless, the SVM's structured approach provided valuable insight into data classification by establishing clear decision boundaries (see **Figure 11**).

Regularization methods: Ridge (L2) and Lasso (L1) regularization methods were tested to address overfitting concerns. Both methods achieved similar accu-

racy levels of 71.81% and sensitivity metrics between 70-72%, indicating that regularization did not significantly improve prediction accuracy. This outcome may be attributed to the prior selection of relevant features in the modeling process (see **Figure 12** and **Figure 13**).

```

Confusion Matrix and Statistics

      Reference
Prediction  0  1
0  108  19
1   28 104

      Accuracy : 0.8185
      95% CI : (0.7661, 0.8635)
      No Information Rate : 0.5251
      P-Value [Acc > NIR] : <2e-16

      Kappa : 0.6374

      McNemar's Test P-Value : 0.2432

      Sensitivity : 0.7941
      Specificity : 0.8455
      Pos Pred Value : 0.8504
      Neg Pred Value : 0.7879
      Prevalence : 0.5251
      Detection Rate : 0.4170
      Detection Prevalence : 0.4903
      Balanced Accuracy : 0.8198

      'Positive' Class : 0

```

Figure 10. Result from random forest model.

```

Confusion Matrix and Statistics

      Reference
Prediction  0  1
0  101  31
1   35  92

      Accuracy : 0.7452
      95% CI : (0.6875, 0.7971)
      No Information Rate : 0.5251
      P-Value [Acc > NIR] : 2.888e-13

      Kappa : 0.4899

      McNemar's Test P-Value : 0.7119

      Sensitivity : 0.7426
      Specificity : 0.7480
      Pos Pred Value : 0.7652
      Neg Pred Value : 0.7244
      Prevalence : 0.5251
      Detection Rate : 0.3900
      Detection Prevalence : 0.5097
      Balanced Accuracy : 0.7453

      'Positive' Class : 0

```

Figure 11. Result from SVM model.

Confusion Matrix and Statistics

```

Reference
Prediction 0 1
0 96 33
1 40 90

Accuracy : 0.7181
95% CI : (0.6591, 0.7721)
No Information Rate : 0.5251
P-Value [Acc > NIR] : 1.701e-10

Kappa : 0.4364

McNemar's Test P-Value : 0.4825

Sensitivity : 0.7059
Specificity : 0.7317
Pos Pred Value : 0.7442
Neg Pred Value : 0.6923
Prevalence : 0.5251
Detection Rate : 0.3707
Detection Prevalence : 0.4981
Balanced Accuracy : 0.7188

'Positive' Class : 0
    
```

Figure 12. Result from L2 model.

Confusion Matrix and Statistics

```

Reference
Prediction 0 1
0 98 35
1 38 88

Accuracy : 0.7181
95% CI : (0.6591, 0.7721)
No Information Rate : 0.5251
P-Value [Acc > NIR] : 1.701e-10

Kappa : 0.4355

McNemar's Test P-Value : 0.8149

Sensitivity : 0.7206
Specificity : 0.7154
Pos Pred Value : 0.7368
Neg Pred Value : 0.6984
Prevalence : 0.5251
Detection Rate : 0.3784
Detection Prevalence : 0.5135
Balanced Accuracy : 0.7180

'Positive' Class : 0
    
```

Figure 13. Result from L1 model.

Feature Importance Analysis: As illustrated in **Figure 14**, feature importance analysis revealed that the most significant predictors were “Runtime,” “IMDb Rating,” and “Rotten Tomatoes Rating,” each with high “MeanDecreaseGini” values. These variables significantly impact the model’s decision-making process. Addi-

tionally, the feature “Has Won Award” demonstrated importance, highlighting the role of award recognition in enhancing a film’s credibility and audience interest.

	MeanDecreaseGini
Has.Won.Award	43.35426
Is.a.Superstar	32.24294
IMDb.Rating	58.92510
Rotten.Tomatoes.Rating	57.58034
Runtime	59.16796

Figure 14. Result from feature importance analysis.

5. Discussion

This study aimed to predict movie success using various machine learning models and identify the key factors contributing to a film’s box office performance and recognition. The findings offer significant insights for the film industry, aligning with and expanding upon existing research.

5.1. Model Performance

The Random Forest Model outperformed other models with an accuracy of 81.85% and a sensitivity of 79.41% for class 0. Its strength lies in its use of multiple decision trees, bootstraping, and random feature selection, which enhance predictive power and reduce feature correlation. This ensemble approach effectively handles complex relationships between variables, making it a valuable tool for stakeholders seeking data-driven decision-making.

The Support Vector Machine (SVM) Model achieved an accuracy of 74.52% and a sensitivity of 74.26%. While slightly less accurate than the Random Forest, SVMs excel at defining clear decision boundaries, providing valuable classification insights. However, ensemble methods like Random Forest demonstrate superior predictive capacity when managing a broader set of features.

Regularization techniques (Ridge L2 and Lasso L1) achieved similar accuracy scores of 71.81% and sensitivities between 70% - 72%. The minimal improvement suggests that the dataset’s pre-selected features were already optimized, rendering additional regularization less impactful.

5.2. Key Influential Factors

These findings have significant implications for stakeholders in the film industry, particularly when planning marketing strategies and making casting decisions. The analysis revealed that key features like “Runtime”, “IMDb Rating”, and “Rotten Tomatoes Rating” were highly influential, as reflected in their “Mean-DecreaseGini” scores. High values for these features suggest that focusing on these elements can enhance the predictive modeling of a movie’s potential success. Additionally, “Has Won Award” underscores the role of award recognition in boosting a film’s credibility and audience interest, suggesting that strategic targeting of

awards can enhance promotional campaigns.

Combining star power (“Is a Superstar”) with positive ratings and optimal runtime maximizes a film’s appeal. These findings emphasize the importance of leveraging established actors and directors to attract audiences and drive financial performance.

5.3. Implications for the Film Industry

These results have several implications for film industry stakeholders:

1) Marketing Strategies: Focus on elements that significantly influence success, such as securing high ratings and targeting award nominations. Tailored, cost-effective marketing strategies are preferred over solely increasing budgets, aligning with the concept of diminishing returns on excessive marketing spend.

2) Casting Decisions: Prioritize hiring superstars and renowned directors, as their presence enhances marketability and success, especially during opening weekends.

3) Production Planning: Optimize film runtime and leverage high ratings to improve box office performance and long-term profitability.

5.4. Comparison with Previous Research

The findings align with [Lash and Zhao’s \(2016\)](#) Star Power Theory, which emphasizes the significant impact of well-known actors on a film’s commercial success. The study confirms that star power attracts larger audiences and enhances financial performance, supporting previous research that established actors and directors increase a film’s marketability and success.

The results also echo [Wisniewsky’s](#) research, which demonstrated that films with larger marketing budgets typically perform better ([Wisniewsky, 2023](#)). However, the study highlights potential diminishing returns on excessive marketing spend, reinforcing the need for strategic and targeted marketing efforts rather than focusing solely on budget size.

Furthermore, He and Hu’s findings on the importance of social media engagement and digital marketing as modern predictors of film performance are indirectly supported. The emphasis on ratings and award recognitions, which often influence online discussions and viewer interest ([He & Hu, 2021](#)), suggesting that targeted digital campaigns and leveraging award credibility are essential components of modern film marketing.

5.5. Broader Implications and Future Research

These findings contribute to the understanding of how machine learning can predict box office outcomes and inform strategic decisions. The strong performance of the Random Forest Model suggests that ensemble learning methods are particularly well-suited for handling the multifaceted nature of film success prediction, where multiple factors interact in complex ways.

However, the study’s applicability may vary across different cultural contexts.

For instance, in markets like China, spectacle-driven films and holiday releases may hold greater appeal, whereas in regions like France, local storytelling and regional star power might be more influential. Future research should explore these predictors across diverse cultural landscapes to provide broader insights for optimizing international film success.

5.6. Limitations

The study faced data extraction issues, resulting in missing information for several films, such as *Avatar: The Way of Water*, *Spider-Man: Far from Home*, and *The Conjuring: Last Rites*. These gaps could affect the dataset's completeness and the predictive models' accuracy, potentially limiting the generalizability of the findings. Future studies should aim for more comprehensive data collection to mitigate these limitations.

Future studies could expand on this research by incorporating data from streaming platforms and analyzing how their metrics influence long-term success. The study's findings provide a foundational understanding of the relationship between movie features and financial success, serving as a guide for industry professionals seeking to optimize their strategies in an ever-changing entertainment landscape.

6. Conclusion

This study offers valuable economic insights into the factors driving film success, with significant implications for marketing, casting, and strategic planning within the film industry. Utilizing various machine learning models, the **Random Forest Model** emerged as the most effective predictor, achieving an accuracy of 81.85%. This highlights the model's capability to manage complex variable relationships, making it a powerful tool for stakeholders aiming to make informed, data-driven decisions.

Key predictors identified include "Runtime," "IMDb Rating," and "Rotten Tomatoes Rating," which were instrumental in forecasting a film's box office performance. Additionally, award recognition ("Has Won Award") and star power ("Is a Superstar") were significant contributors, underscoring the importance of leveraging these elements in strategic marketing and casting decisions.

The findings also suggest that while increased marketing budgets can enhance performance, there is a threshold beyond which additional spending yields diminishing returns. This emphasizes the need for targeted and cost-efficient marketing strategies rather than indiscriminate budget increases.

Despite these insights, the study faced limitations, including incomplete data for certain high-profile films, which may affect the generalizability of the results. Future research should aim to incorporate more comprehensive datasets and explore emerging trends, such as the impact of streaming platforms, to provide a more holistic understanding of film success in the evolving entertainment landscape.

In summary, the study demonstrates that integrating critical factors like strong ratings, strategic casting, and optimized marketing can significantly enhance a film's success. These findings advocate for a balanced, informed approach to decision-making in the film industry, leveraging advanced predictive models to navigate the complexities of market dynamics.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Carollo, L. (2024). *U.S.: Film Production & Distribution Revenue 2022*. Statista. <https://www.statista.com/statistics/184140/estimated-revenue-of-us-motion-picture-and-video-industry-since-2005/>
- Chisholm, D. C., Fernández-Blanco, V., Abraham Ravid, S., & David Walls, W. (2014). Economics of Motion Pictures: The State of the Art. *Journal of Cultural Economics*, 39, 1-13. <https://link.springer.com/article/10.1007/s10824-014-9234-1>
- Domestic Box Office for 2023*. Box Office Mojo. <https://www.boxofficemojo.com/year/2023/>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson.
- He, Q., & Hu, B. (2021). Research on the Influencing Factors of Film Consumption and Box Office Forecast in the Digital Era: Based on the Perspective of Machine Learning and Model Integration. *Wireless Communications and Mobile Computing*, 2021, Article ID: 6094924. <https://doi.org/10.1155/2021/6094924>
- Lash, M. T., & Zhao, K. (2016). *Early Predictions of Movie Success: The Who, What, and When of Profitability*. <https://arxiv.org/abs/1506.05382>
- Mitchellmarino (n.d.). *Mitchellmarino/Movie-Profitability-Classifer*. GitHub. <https://github.com/mitchellmarino/Movie-Profitability-Classifer>
- Nichols, B. (2018). *Taking Note: The Film Industry's Contributions to National and State Economies*. National Endowment for the Arts. <https://www.arts.gov/stories/blog/2018/taking-note-film-industrys-contributions-national-and-state-economies>
- O'Brien, R. M. (2007). *A Caution Regarding Rules of Thumb for Variance Inflation Factors*. *Quality & Quantity*, 41, 673-690. <https://doi.org/10.1007/s11135-006-9018-6>
- Scott, G. (2019). Determinants of Box Office Performance: Return of the Regressions. *Major Themes in Economics*, 21, 71-83. <https://scholarworks.uni.edu/mtie/vol21/iss1/7/>
- Skye, N. (2024). *Social Media's Influence on Modern Movie Marketing*. Journotalk. <https://journotalk.com/social-medias-influence-on-modern-movie-marketing/>
- Wisniewsky, Z. (2023). *Breaking down the Box Office: An Analysis of Film Profitability Factors*. UCONN Library. https://digitalcommons.lib.uconn.edu/cgi/viewcontent.cgi?article=1968&context=srhonors_theses