IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

IJARSCT

Impact Factor: 7.301

Volume 3, Issue 12, May 2023

IMDB Movie Rating Prediction Using Machine Learning

Prof. Shubhangi Patil[1], Prof. Jogendrasingh Solanki[2], Govind Majage[3], Nikhil Dhongade[4], Pradosh Kaurase[5], Rohan Mahadik[6]

Professor, Department of Computer Technology[1,2]
Students, Department of Computer Technology[3,4,5,6]
Sinhgad Institute of Technology, Lonavala, Maharashtra, India

Abstract*: Special selection techniques such as correlation analysis and regression feature elimination are used to identify the most valuable features. Various machine learning algorithms, including linear regression, decision trees, random forest, and gradient boosting, have been used to develop predictive models. Performance is measured using metrics such as mean squared error (MSE), mean squared error (RMSE), and R-squared. Experimental results showing the effectiveness of the proposed models in IMDb movie prediction. This model provides a high level of accuracy in rating, taking into account certain characteristics such as genre, director, actors, production budget, release date and user reviews. These findings provide better insights into the factors that affect film performance and help filmmakers and audiences make more informed decisions. This research contributes to the development of video prediction using machine learning. Design provides useful information to the film industry, helping to select films and optimize the decision-making process. The findings could improve the movie experience for viewers and boost the overall performance of the entertainment industry.*

Keywords: Machine learning, Decision tree classifier, movie prediction, IMDB

I. INTRODUCTION

Movie ratings play an important role in the movie industry, influencing audiences, box office success and critical response. The ability to predict movie ratings is important to filmmakers, distributors, and audiences alike. By understanding the factors that influence filmmaking, marketing professionals can make informed decisions about production, marketing and distribution strategies. This research paper presents a machine learning-based approach to estimating IMDb movie ratings using large amounts of available data and using advanced algorithms for accurate prediction.

The IMDb (Internet Movie Database) platform is a popular online site that provides detailed information about movies, including user-generated ratings and reviews.

IMDb ratings are based on a scale of 1 to 10 and represent the average user rating of a movie. Estimating the IMDb score is a difficult task, as there are many factors that influence the audience, such as the genre of the movie, actors, directors, production budget, release date, and user reviews. Machine learning techniques show promise for this tricky prediction problem by revealing patterns and relationships in data.

The aim of this study is to develop a prediction model that accurately predicts IMDb movie ratings based on relevant features. To achieve this goal, an extensive database of historical video data and related metrics was used.

II. METHODOLOGY

- Data Scraping: The first step in this research is to collect detailed data including historical movie data and their corresponding IMDb scores. The information should include various movie details such as genre, cast, director, production budget, release date, and user reviews. IMDb has a large collection of movie information and ratings, making it a useful source of information.
- Preliminary Data: Preliminary steps are taken to ensure the quality of the data and its suitability for machine learning models while retrieving data. This includes handling of missing values, which may include

DOI: 10.48175/568

assignment procedures or removal of missing data. Numerical features are normalized to ensure they are in balance, while categorical variables are encoded into numeric representations using techniques such as one-bit encoding or label encoding.

- Feature Selection: Feature selection is an important step in identifying key features associated with IMDb movie ratings. Each algorithm has advantages and disadvantages, and the choice depends on factors such as dataset size, feature complexity, and interpretability.

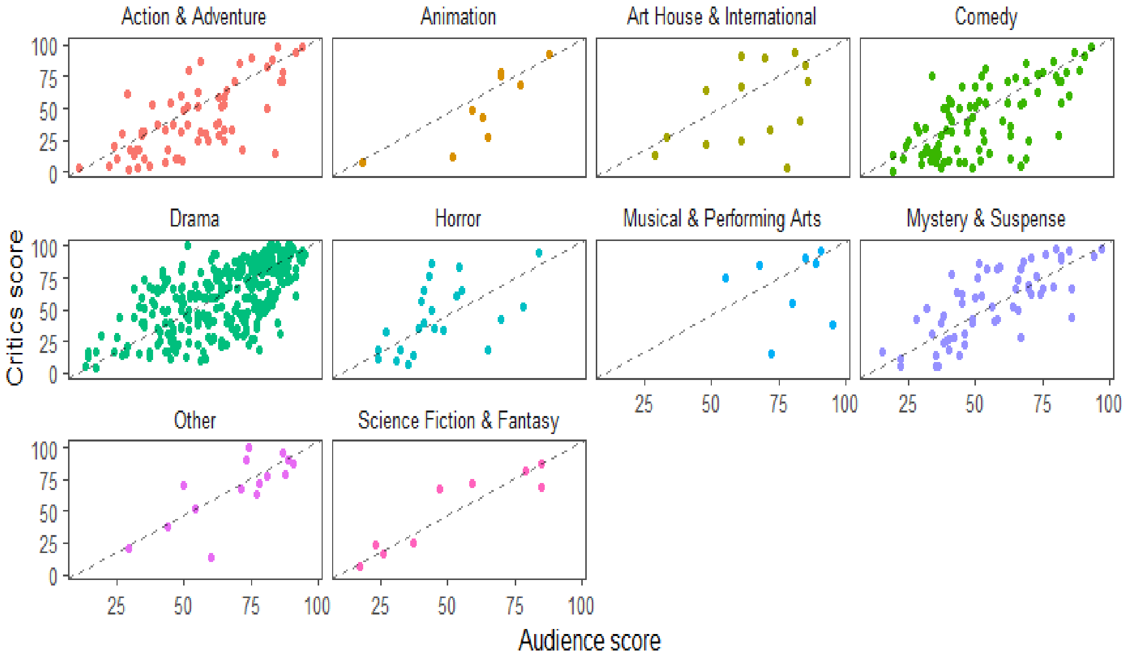


Fig. Feature Selection

- Model Training and Evaluation: Use prepared data to train selected machine learning models. The data is divided into training and testing subsets to test the performance of the model. During training, the model learns the relationship between input and IMDb scores. This process helps to refine the model and improve prediction accuracy.

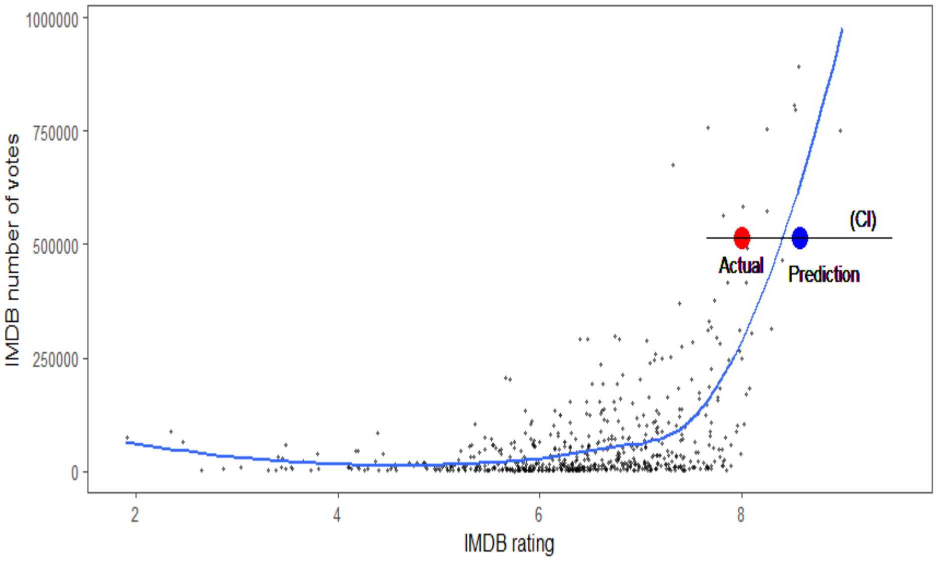


Fig. Model Evaluation

- Model distribution and estimation: Once the best performance model is determined, it can be sent to predict IMDb videos. A new, invisible video can be fed to the training model that will display a rating based on its characteristics. Such forecasts can help filmmakers, distributors, and viewers make decisions about movie selection, marketing strategies, and audiences.
- Performance Analysis and Interpretation: After the estimated score is obtained, the performance of the model can be further analyzed and explained. The strongest features can be analyzed to understand their impact on metrics. Additionally, the strengths and limitations of the model can be discussed and insights can be gained about the factors that contribute to IMDb movie ratings.

III. WORKING

IMDb movie prediction machine learning involves training a model to predict movie ratings based on various features or characteristics. We have used Sklearn decision Tree Classifier algorithm for this, Below is an overview of IMDb's steps to create machine learning based movie ratings:

- Data Collection: Collecting data containing Movie information, including attributes such as title, genre, director, actors, release year, show time. , budget and other relevant information. You can get this information from IMDb or other movie sources.
- Data pre-processing: Cleans and pre-process the collected data to ensure it is in a format suitable for training machine learning models. This includes handling missing values, converting categorical variables to numeric representations (eg.one-bit encoding or tag encoding) and scaling number properties
- Feature Engineering: Extracting key features from existing data to help improve model performance. For example, you can create new features such as the number of actors in the movie, the average rating of the director's previous movies, or the available credits listed according to past IMDb scores.

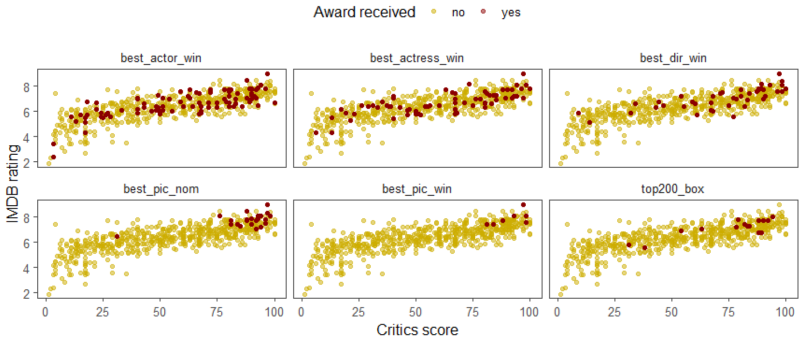
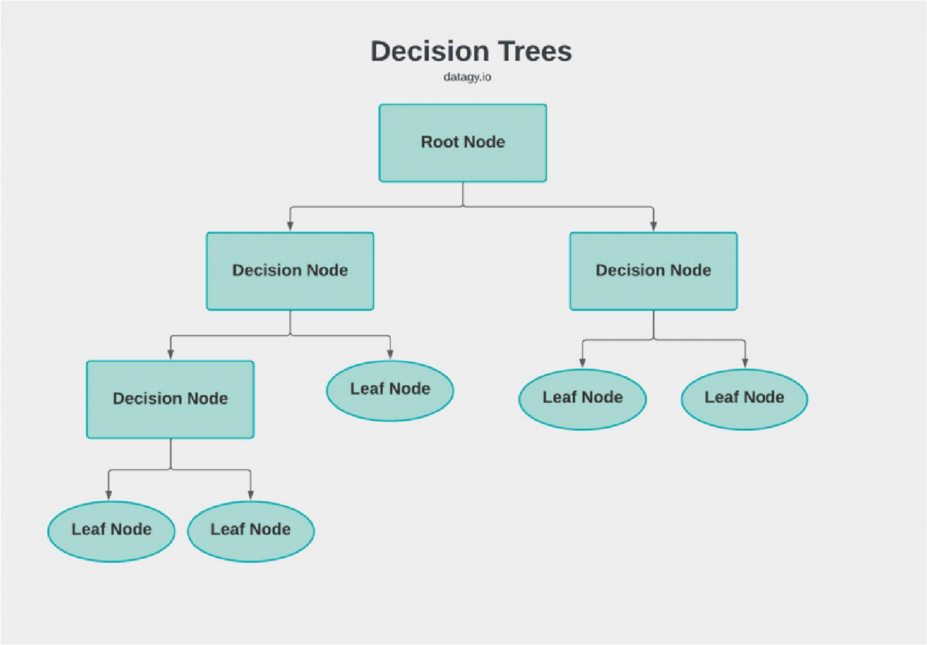


Fig. Feature Engineering

- Split Dataset: Split data into training and testing subsets. The training process will be used to train the machine learning model, while the testing process will be used to evaluate its performance.
- Model Selection: Select the appropriate Machine Learning Algorithm to predict movie prices. Some common algorithms for such regression tasks include linear regression, decision trees, random forests, and gradient boosting algorithms.
- Model Training: Trains selected machine learning models using training data. In this step, the model learns the underlying structure and the relationship between input features and video playback.
- Model Evaluation: Evaluate the performance of training models using test data. Common statistical measures for regression functions include mean square error (MSE), root mean square error (RMSE), and R-squared.
- Model fine-tuning: If the performance of the model is not satisfactory, you can tune the algorithm's hyperparameters or try different algorithms to improve the prediction. This step involves experimenting with different configurations to optimize the model's performance.
- Model Deployment: Once you are satisfied with the model's performance, you can use it to make predictions about new video data. The distribution model can be used to predict the IMDb score of a movie based on its characteristics.



- Fig. Decision Tree Classifier Algorithm

IV. LITERATURE SURVEY

There is a lot of work on this project. The massive focus on analyzing and predicting movie success became so popular in 2006 when Netflix announced a nearly $1 million bonus for the top team that created the video with its video rating algorithm.

Google has apps for search volume for movie trailers. After the search volume is calculated and analyzed, it can be estimated that holiday revenue will open up to new ads.

Studies of trendsetters (clark & an a-2017, 2017, social media, social media, user preferences features (Oghina et al., 2012, El Assay et al., 2013 jager et al., 2013, Yaffen Lu and Maciejewski, 2013) used and requested by other groups before taking the group as a group, followed by the film's internal features and all He divided the videos into two groups to analyze the factors affecting their properties.

The first group includes actors, genres, directors, etc., who determine the film itself and have an impact on the quality and effectiveness of the film.

Awad, Delarocas, and Zhang (2004) reviewed the literature on videogames for videogames. They developed a model based on video reviews to predict revenue, and also analyzed the relationship between professional communication and customer service, traditional and online word of mouth. Finally, they concluded that professional expertise, traditional customer communication, and online word of mouth had a positive effect on video performance.

V. CONCLUSION

Many different approaches to this problem are used with different methods and functions. The data used were collected from the publicly available IMDb database. The results show that the implicit semantic analysis achieves the best estimate. Future plans include more detailed testing with different materials, as currently used materials do not differ from user ratings.

Some systems have not been adequately tested due to their computational complexity and need to be thoroughly tested. Other methods suggested in related studies should also be considered. A combination of individual items and the use of voting weights can be tested for predictive purposes. Finally, the development of a new video rating prediction algorithm is a future option.

DOI: 10.48175/568

REFERENCES

[1]. Internet Movie Data Base. URL: https://www.imdb.com

[2]. M. H. Latif and H. Afzal, “Prediction of movies popularity using machine learning techniques,” International Journal of Computer Science and Network Security (IJCSNS), vol. 16, no. 8, p. 127, 2016.

[3]. http://www.statista.com/statistics/259985/global-filmedentertainment-revenue/

[4]. https://www.researchgate.net/publication/222530390

[5]. E. Frank, I.H. Witten, Data Mining, Morgan Kaufmann Publishers, 2000.

[6]. R. Parimi and D. Caragea, “Pre-release box-office success prediction for motion pictures,” in International Workshop on Machine Learning and Data Mining in Pattern Recognition. Springer, 2013, pp. 571–585.

[7]. N. Quader, M. O. Gani, D. Chaki and M. H. Ali, "A machine learning approach to predict movie box-office success," 2017 20th International Conference of Computer and Information Technology (ICCIT), Dhaka, 2017, pp. 1-7, doi: 10.1109/ICCITECHN.2017.8281839.

[8]. Oghina, M. Breuss, M. Tsagkias, and M. De Rijke, “Predicting imdb movie ratings using social media,” in European Conference on Information Retrieval. Springer, 2012, pp. 503–507.