JSRT JOURNAL

PAGES:1-4
1/3/24

JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY(JSRT) VOLUME-2 ISSUE-1 JANUARY
Registered under MSME Government of India ISSN: 2583-8660

Movie Success and Rating Prediction Using Data Mining

Ambresh Bhadrashetty[1], Surekha Patil[2]

[1]*Assistant Professor, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University Kalaburagi, Karnataka, India.* ambresh.bhadrashetty@gmail.com

[2]*Post Graduate Student, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University Kalaburagi, Karnataka, India* surekhapatil1062000@gmail.com

ABSTRACT

A large sum of money is invested annually in the production of several films. The primary objective of our study is to predict, using specific credits (both static and social/dynamic), if the film will be a true hit or a complete bust. There are a lot of set aspects that impact a movie's success or failure, such as the kind, budget, entertainers, chief, creator, creation home, delivery date, and so on. Looking at the film through the lens of online entertainment, we would search for dynamic hashtags that are now trending on Twitter. It is critical to find the attribute relationships and use an information mining calculation to get the result. To find out whether any movie is lucky or not, we apply all the information mining tools. This method is very helpful for those who build things. due to the fact that it gives them the opportunity to review films before to their release, which greatly influences their self-presentation and enhances their outcomes.

Keywords: IMDB dataset, Data mining, Machine Learning(ML), Movies, Prediction

I. INTRODUCTION

Movies play a significant role in the entertainment industry's bottom line and are critically acclaimed works of art. Studio heads, producers, and distributors must, therefore, accurately predict a film's box office take and critical reception before to its release. The advent of data mining as a predictive analytics tool has made it possible to sift through mountains of film-related data in search of useful insights.

Data mining is the process of investigating large datasets for relevant patterns, correlations, and trends. Several factors impact a film's success and critical acclaim, and data mining tools may help investigate these factors. First, the project will provide professionals in the film industry with a tool to help them make decisions about production, marketing, and distribution. Second, data mining will get a boost as this project shows how advanced techniques can be used to predict how well a movie will do at the box office. More specific information on the data used for this project will be provided in the parts that follow.

II. RELATED WORKS

Acquiring subject-specific knowledge is the goal of artificial intelligence. The accumulation of knowledge include understanding of dark energy (74% of the cosmos), dark matter (22%), and visible matter (4%). Both biological and non-biological sensors, such as robots, televisions, mobile phones, cameras, microscopes, radar, computers, etc., may be used to collect this information [1]. Tools that are increasingly complicated and advanced are needed to handle the ever-increasing amounts of data in contemporary research and industry. There is a continual need for innovative methods and tools to help us convert enormous data into valuable information and understanding, even while data mining technology has made massive data collecting much simpler. Significant progress has been achieved in the realm of data mining since the release of the last version [2]. One of the most prominent entertainment industries is the film industry. Although the Nigerian film industry cranks out a plethora of films for the general people to enjoy, only a select few manage to become commercial successes. The advent of success prediction for movies has the potential to revolutionize the business by assisting filmmakers and producers in making more informed choices that ultimately benefit the bottom line. The authors of this paper suggest a methodology for predicting the box office performance of forthcoming films using data mining and machine learning algorithms, with the help of certain predetermined factors. In order to forecast a film's financial success or failure, several factors must be considered. These include the dataset required for data mining, which includes information about the film's writers, directors, actors, and actresses as well as the film's marketing and production budget, target demographic, theater location, film's release date, and

the performances of competing films that share the same date. Before purchasing a movie ticket, this model also assists moviegoers in determining if an upcoming film is a blockbuster, hit, or has a high success rating. Internet Movie Database MetaData, after undergoing an initial cleansing and integration procedure, was subjected to data mining algorithms [3]. Although networking research has long made use of machine learning and AI, much of this work has been on supervised learning. Unsupervised machine learning has become more popular as of late for enhancing network performance and offering services like traffic engineering, anomaly detection, Internet traffic categorization, and quality of service optimization utilizing unstructured raw network data [4]. Lorene Wales's The Complete Guide to Film and Digital Production, now in its third edition with extensive updates and revisions, provides a thorough overview of the many jobs and responsibilities involved in making films and digital videos, from ideation and pre-production through distribution and marketing. With a variety of example checklists, timetables, accounting paperwork, and downloadable forms and templates, Lorene Wales provides a practical strategy that is ideal for projects of any budget and scope. She explains every step and crucial role/position in a film's existence and gives a hands-on approach. In addition, we cover the most recent production-related mobile applications, tax incentives, the DIT, set safety, and a more in-depth discussion of copyright, fair use, and other legal issues in a newly extended chapter.

Documents such as schedules, accounting paperwork, releases, and production checklists are available for download on the accompanying website, which also features video tutorials, a personnel hierarchy, a guide to mobile apps that can be helpful during production, and PowerPoints that instructors can use. [5].

III. METHODOLOGY USED

Random Forest Algorithm: This method builds decision trees using randomly selected dataset samples. Once all of the decision trees have had their outcomes voted on, the final forecast is created.

Naïve Bayer's Algorithm: Naive Bayes predicts the likelihood of different classes supporting different qualities using a same procedure. Text categorization is the typical use of this method. Using the comments on YouTube trailers and teasers, we trained this algorithm to forecast a film's projected rating.

Depending on the dataset and task, one approach may be more appropriate than the other; each has advantages and disadvantages. To find the best algorithm for a certain movie prediction job in terms of accuracy and precision, it's necessary to test and compare many algorithms. Predicting the likelihood of success or failure for future films is the overall goal of the created model, which employs data mining methods, suitable algorithm selection, and historical data. Nevertheless, it should be mentioned that there are still a lot of unknowns in the film business that might affect how accurate the projections are.

Classification techniques: Such technqiues to aid the created model in foretelling the future box office performance of films. The model was built using the correct machine learning approach utilizing 80% of the dataset's data. With the remaining data set, the model is trained. The proposed approach relies on classification methods to forecast the success of upcoming movies.

IV. RESULTS AND DISCUSSION

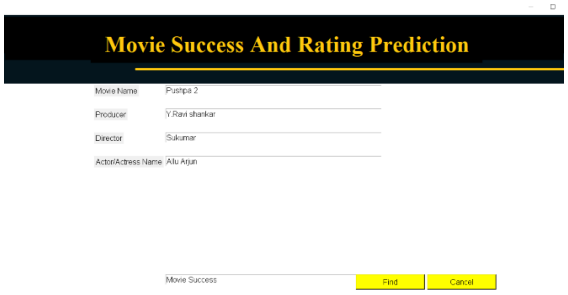

Fig 1: Prediction Window

Details such as the film's title, director, producer, actors, actresses, etc. are shown in the prediction box. These things have a role in determining the future movie's success or failure. In addition, it will provide the movie's rating (out of 10).

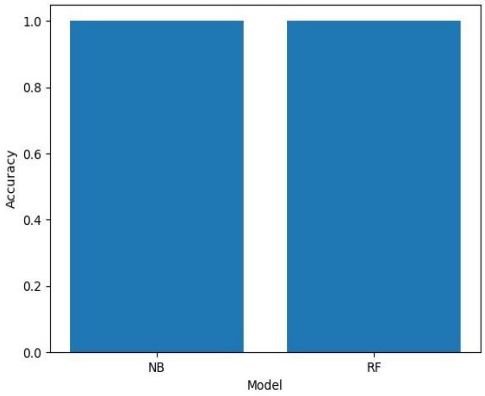

Fig 2: Accuracy Graph

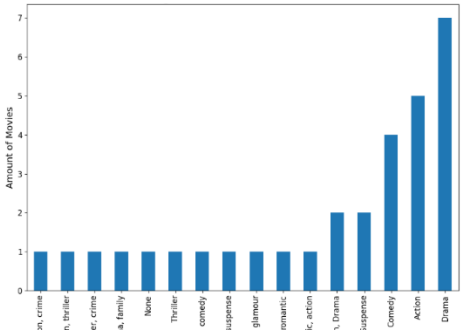

Fig 2: Genre vs Amount of Movies

V. CONCLUSION

After much deliberation, we have come to the conclusion that it is not possible to forecast a new film's anticipated rating based on the comments left on trailers and teasers on YouTube prior to its actual release. The YouTube API pulls out more negative than good comments since the ones with the most remarks show up first. More negative than positive results appear in the model's output on a regular basis. This means that our model isn't always accurate. Although our tried-and-true method works, we can't foretell how a new picture will do in theaters before it opens.

We therefore came to the conclusion that it was possible to forecast a new film's box office performance prior to its debut by analyzing internet data and features. We limited the number of characteristics to 5 to make it easier to handle the large quantity of data and make sure that only the relevant information is utilized and understood by the user. We trained the algorithm to predict the ticket sales of upcoming movies. The model achieved an accuracy level of over 70% across all test sets. So, we concluded that our system is doing its job well and can predict a film's box office performance before its debut.

REFERENCES

[1] Grewal, D. S. (2014). A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering. IOSR Journal of Computer Engineering (IOSR-JCE), 16(2), 9-13.
[2] Han, J., Kamber, M. & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). New York: Morgan Kaufmann. Kumar, V., Ramakrishnan, G., & Li, Y. (2018). A framework for automatic question generation from text using deep reinforcement learning. ArXiv, abs/1808.04961. McCarthy, J. (2000). "Review of Artificial intelligence: A General Survey." [Online]. Available: http://wwwformal.stanford.edu/jmc/reviews/lighthill/lighthill.html.
[3] Devansh Priye, & Sumit Sangwan. (2023). A Study of Students Stock Market Participation and Awareness. Journal of Scientific Research and Technology, 1(8), 70–90. https://doi.org/10.61808/jsrt73
[4] Sabina Anjum, & Asra Fatima. (2023). Predictive Analytics For FIFA Player Prices: An ML Approach. Journal of Scientific Research and Technology, 1(6), 204–212.
[5] Dr. Shubhangi D C, Dr. Baswaraj Gadgay, & S. Anita. (2023). Leverage Machine Learning To Infer Proof of the Nipah Influenza. Journal of Scientific Research and Technology, 1(9), 13–20. https://doi.org/10.61808/jsrt75
[6] Nithin, V.R, Pranav, M., Badu, P. B., &Lijiya, A. (2014). Predicting movie success based on IMDB data. International Journal of Business Intelligents, 3(2), 34-36. 10.20894/IJBI.105.003.002.004. 43
[7] Ogunleye, F. (2004). A Report from the Front: The Nigerian Videofilm. Quarterly Review of Film and Video, 21(2), 79-88. DOI: 10.1080/10509200490272991.
[8] Paul Arunkumar J, Dr. K. Subathra, Dr. S. Senthilkumar, Prabhavathy R, & Rohan Thomas Jinu. (2023). Leveraging the power of social proof on online consumer behaviour. Journal of Scientific Research and Technology, 1(5), 31–39.
[9] Mohammed Maaz, Md Akif Ahmed, Md Maqsood, & Dr Shridevi Soma. (2023). Development Of Service Deployment Models In Private Cloud. Journal of Scientific Research and Technology, 1(9), 1–12. https://doi.org/10.61808/jsrt74
[10] Wolpert, D. H. & Macready, W. G. (1997). "No free lunch theorems for optimization," in IEEE Transactions on Evolutionary Computation, 1(1), 67-82. doi: 10.1109/4235.585893.
[11] Divya Kalra, Sanjeev Sharma, & Aayush Patel. (2023). A Review on Impact of Digital Marketing on Consumer Purchase Behaviour. Journal of Scientific Research and Technology, 1(3), 15–20.
[12] Dr. Rekha J Patil, Indira Mulage, & Nishant Patil. (2023). Smart Agriculture Using IoT and Machine Learning. Journal of Scientific Research and Technology, 1(3), 47–59.
[13] Morgan Kaufmann. Kumar, V., Ramakrishnan, G., & Li, Y. (2018). A framework for automatic question generation from text using deep reinforcement learning. ArXiv, abs/1808.04961.
[14] Mesakar, S. & Chaudhari, M. (2013). A Review of Clustering Algorithms. International Journal of Computer Science and Technology (IJCST), 4(1), 255-257.
[15] Smt. Jayanti K, Ravi Pare, Saurabh S P, & Shashank S H. (2023). Exploratory Analysis Of Geo-Location Data. Journal of Scientific Research and Technology, 1(3), 60–67.