



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 5, Issue 2)

Available online at: www.ijariit.com

A survey on motion picture prediction using data mining

Abhishek More

a.more3111997@gmail.com

K.C. College of Engineering,
Mumbai, Maharashtra

Sharique Khan

khansharique69@gmail.com

K.C. College of Engineering,
Mumbai, Maharashtra

Thabreez Izzudeen

thabreezizzudeen@gmail.com

K.C. College of Engineering,
Mumbai, Maharashtra

ABSTRACT

In this study, a mathematical model used to predict an upcoming movie success class. This will help to suggest as your interest before release based on several attributes and it will take a suggestion from different linked social sites to determine the interest of the audience. It also shows the most interesting result on top of your list. Many film studios collectively produce several hundred movies every year; the budget of this movie is of the order of hundreds of millions, making their Box office success absolutely essential for the survival of the Industry. Knowing which movie is going to succeed and which are going to fail before the release could benefit the Industry. In this paper, we have enlisted some existing movie prediction using data mining for predicting success and failure in considering their advantages and disadvantages.

Keywords— Data mining, Predictive analysis, YouTube, Twitter, IMDb

1. INTRODUCTION

We live in a generation where a movie releases every week. In movies, pictures in motion sound effects like 2D, 3D and much more are used. As a result, a large amount of data called big data is generated from each movie. As an example, let's assume you are shooting a 4k movie for two hours which approximately have 20 hours of raw footage so that will be about 2.5 GB of storage needed just for raw footage while transcoding you need to recalculate for the new files, render files and output files. So in this scenario, it will be comfortable with 6TB of storage with a 3TB drive for the backup of raw footage.

All data is not useful some are chaotic data which are just noise and some are well structured and informed data. By the advent of data analytics, the prediction process has been made intelligent by considering past datasets and applying different data analytical techniques to predict the result.

Big data refers to a huge amount of dataset (Exabyte or petabyte) which makes it difficult to handle such a large amount of data which is growing at vast speed. This dataset is not easy to handle by traditional tools, so data mining tools are used.

2. DATA MINING

It is the process of finding different patterns and algorithms within large datasets to predict the result. Data mining is a versatile subfield of Information Technology with a goal to extract information from datasets and transform the information into a coherent structure for further usage. By the time passing by we are seeing massive numbers of data produced every year. Disorganized data alone makes up 90% of the digital universe. Big data analyzing help us to Shift all the disorganized and repeated noise from the big dataset. And Understanding of significant data to make good use of it and help in Accelerating in making an enlightened decision.

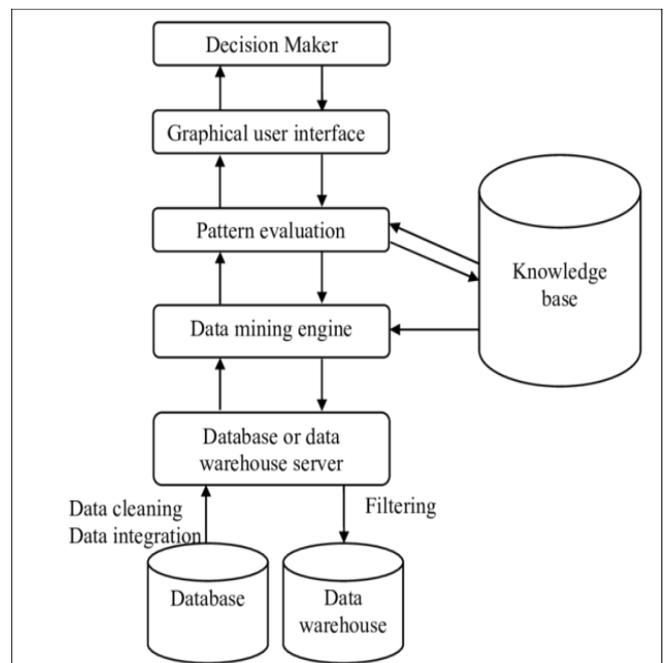


Fig. 1: Big data architecture

There square measure two types of information analysis which will be used for extracting models describing necessary categories or to predict future information trends. These are as follows:

- Classification
- Prediction

Classification model forecast categorical class labels and prediction models forecast continuous valued function. For e.g. Classification model to categorize bank loan applications as either safe or risky and designing a Prediction mode to predict the monthly budget of customers depending on their income and occupation.

2.1 Classification

It may be a data processing technique that assigns classes to a group of data so as to help in additional correct predictions and analysis. For e.g. a large amount of data are generating in today's world of "big data". A database with multiple types of data – a terabyte is one trillion bytes of data. Just give it a thought Facebook itself generate 600 terabytes of new data in a single day. To deal with this amount of data some automatic methods are extracted to make useful information has been developed and Classification is one of them.

2.2 Prediction

In data mining, it is used to spot information points strictly on the outline of connected information worth. It's not essentially associated with future events however the used variables are unknown. Prediction derives the connection between a factor you recognize and a factor you would like to predict for future reference. For example, prediction models in the data mining area unit are utilized by a business manager who predicts that how much quantity a particular client can pay throughout an acquisition, so the upcoming sale quantity is planned consequently. The prediction in data processing is considered a Numeric Prediction. Typically, multivariate analysis is employed for prediction. For prediction of data, some algorithms are used such as C4.5, k-means, Support vector machines, Apriori, EM, PageRank, AdaBoost, KNN, Naive Bayes and CART.

3. LITERATURE REVIEW

In paper [1], they're talking about a prediction engine that they have enforced that uses classification and fuzzy logic to categorize a movie as successful or not, additionally as using out their own formula to calculate a target variable which is the IMDB score of a movie based on numerous parameters. Fuzzy Logic: Developing and maintaining an understanding of technical subjects across a broad range of areas means you may have a lot of data and skills to attack issues when they appear and formulate innovative solutions to those issues.

In paper [2], here they report on an attempt to create a minimalistic predictive model for the financial success of movies based on collective activity information of online users. They show that the recognition of a movie is often predicted much before its release by measuring and analyzing the activity level of editors and viewers of the corresponding entry to the movie in Wikipedia (Encyclopedia). In order to find the corresponding articles in Wikipedia, we use the category system of Wikipedia. Wikipedia articles are classified into one or additional classes by users. We tend to match the title of the films within the mojo database with the title of Wikipedia pages in categories films.

In paper [3], in their study, the mathematical model developed to predict the success and failure of the upcoming movies concerned with finding a correlation between numerous attributes using X^2 analysis. Correlation is a measure of dependence between 2 variables. The correlation may be negative or positive. A positive correlation indicated that the 2 variables increase or decrease in parallel, whereas, a negative correlation indicated that the two variables change in opposite directions.

In paper [4], they have applied data mining tools to generate fascinating patterns for predicting box office performance of movies using information collected from multiple social media and internet sources including Twitter, YouTube, and the IMDb movie database. There are 3 steps in their experiments: normalizing the Training information, applying K-Means clustering and generating a predictive model. For normalization, they tend to apply the subsequent straightforward min-max method, followed by applying a spherical function to map every attribute value into an integral rank of one to ten. For clustering, they used the K-Means tool from weka.

In paper [5] In their analysis work, they have developed a mathematical model for predicting the success of the films in terms of the revenue model and rating model. They have developed a model in which the past information record of every element [e.g. actor, actress, genre, director] that influences the success or failure of a movie is considered. Knowing which movies are seemingly to succeed and which are seemingly to fail before the release may benefit the production organizations greatly because it will alter them to focus their advertising campaigns.

Basically, they tried to use the supervised Learning mechanism to the device on the premise of their main aim of the project i.e. to come up with an estimation of the pre-release sales revenue of a Bollywood movie.

In paper [6], machine learning algorithms are used for predictive analysis. Machine learning algorithms applied to a standard, collected from movies databases, and social media options (text comments on Tweets, YouTube). Mining the attributes and contents of social media provides them with a chance to get social structure characteristics, analyze action patterns qualitatively and quantitatively, and generally the power to predict future human-related events. The results of this project predict the success with management and use of opinions from social media and different social media options. This paper presents the comparison of typical options with Social Media options in determining the recognition of films.

4. PROPOSED SYSTEM

In this study, we'll assemble the data from multiple social sites and implement our algorithm to predict the types of trailers that the user is following on social sites as well as his/her tweets, search history, comments and sharing post on Facebook. With this data, we can foretell the taste of a user. On the base of these attributes and data generated by movies is analyzed and using our algorithm it'll predict the upcoming movies in which the user is interested and will notify the user. Moreover, it will recommend the user to book the tickets.

5. DIAGRAM

Proposed system diagram is shown in figure 2.

6. CONCLUSION

In this paper, we have surveyed different approaches for the movie's prediction. Our signature algorithm is tending to predict the movie's success with ease, however still compelling enough to create smart predictions. Compared with different proposed strategies from the challenge it's by far less refined, though its strengths consist its simplicity. So, by using our prediction engine, users will evaluate beforehand if the movie is worth watching and consequently make their choices.

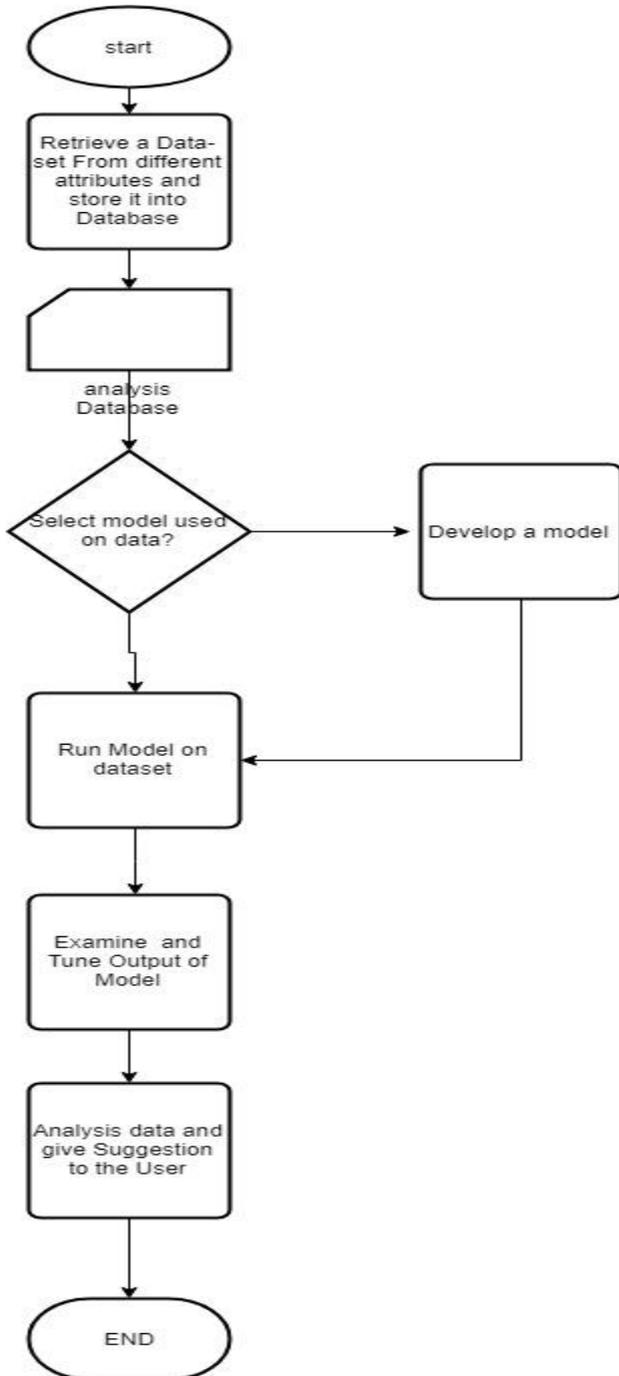


Fig. 2: Proposed system block diagram

7. REFERENCES

- [1] Sanjai Pramod, Abhisht Joshi, Geetha Mary. A” Prediction of Movie Success for Real World Movie Data Sets”, Prediction of Movie Success for Real World Movie Data Sets.
- [2] Ma’rton Mestya’n1, Taha Yasseri1, 2, 3*, Ja’nos Kerte’sz1, 3, 4.” Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data”, Received November 5, 2012; Accepted June 28, 2013; Published August 21, 2013
- [3] Javaria Ahmad*, Prakash Duraisamy*, Amr Yousef†, Bill Buckles, “Movie Success Prediction Using Data Mining”, 8th ICCCNT 2017 July 3 -5, 2017, IIT Delhi, Delhi, India.
- [4] Krushikanth R. Apala, Merin Jose, Supreme Motnam, C.-C. Chan, Kathy J. Liszka, and Federico de Gregorio1, “Prediction of Movies Box Office Performance Using Social Media”, ASONAM’13, August 25-29, 2013, Niagara, Ontario, CAN.
- [5] Harsh Taneja1, Anupam Dewan2, Vineet Bhardhwaj3, “Pre-Release Success Quotient Prediction of Movies”, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391
- [6] Babita M. Jangid, Chaitali K. Jadhav, Swati M. Dhokate, Grish M.Jadhav4, Prof. G.M.Bhandari5,” Survey on Movies Popularity Prediction System Using Social Media Feature”, (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 9, September 2016.