

# SCRUTINIZATION OF MOVIE SUCCESS USING MACHINE LEARNING TECHNIQUE

Dr.M.Rajeswari  
Assistant Professor,  
Department of B.Com (Business Analytics),  
PSGR Krishnammal College for Women, Coimbatore, India.  
[rajeshwarim@psgrkcw.ac.in](mailto:rajeshwarim@psgrkcw.ac.in)

S.Sowndarya<sup>2</sup>  
UG Scholar,  
Department of B.Com (Business Analytics),  
PSGR Krishnammal College for Women, Coimbatore, India.  
[sowndarya447@gmail.com](mailto:sowndarya447@gmail.com)

**ABSTRACT:** The classical factors include producer, production house, director, cast, runtime of the movie, the genre, the script, time of release etc.,. All these factors play a very vital role in determining whether a film will be successful. Polynomial regression means plotting up a graph and it shows the relationship between dependent variable y and independent variables x. This project helps to analyse the success of the film using the machine learning algorithm. Data Science paves a way to extract useful information and to gain insight on it.

**Keywords** – Machine Learning , Data Science, Polynomial Regression, Ratings, Predictive analysis

## [2] I.INTRODUCTION

A film, also called a movie, motion picture or moving picture, is a work of visual art used to simulate experiences that communicate ideas, stories, perceptions, feelings, beauty, or atmosphere through the use of moving images. In 2010, Asur and Huberman worked on predicting outcomes based on social media content [2]

In the United States of America, 1000s of films are released every year. Since the 1920s, the American film industry has grossed more money every year than that of any other country [4].The success prediction of a movie plays a vital role in movie industry because it involves huge investments. However, success cannot be predicted based on a particular attribute.

Number of movies are released every week. There is a large amount of data related to the movies is available over the internet, because of that much data available, it is an interesting data mining topic. In 2015, Lash and Zhao proposed a way to predict about movie investments [3]. The prediction of movies is complex problem. Every viewer, producer, director's production houses all are curious about the movies that how it will perform in the theatre. Many work has been done relating to movies using social networking, blogs articles but much less has been explored by the data and attributes related to a movie which is continuous and in different dimensions. We have used IMDB for our experimentation. We created dataset and then transformed it and applied machine learning approaches to build efficient models that can predict the movies popularity [8].

The proposed research provides a quite efficient approach to predict IMDb score on IMDb Movie Dataset.

## [3] II. OBJECTIVE

To analyse the factors determining the success of the film using one of the machine learning techniques.

## [4] III. RELATED WORK

Predictive Analytics is a statistical method that utilizes algorithms and machine learning to identify trends in data and predict future behaviors [10].

With increasing pressure to show a return on investment (ROI) for implementing learning analytics, it is no longer enough for a business to simply show how learners performed or how they interacted with learning content. It is now desirable to go beyond descriptive analytics and gain insight into whether training initiatives are working and how they can be improved [6]. Predictive Analytics can take both past and current data and offer predictions of what could happen in the future. This identification of possible risks or opportunities enables businesses to take actionable intervention in order to improve future learning initiatives.

Dataset is mixed with both nominal and numeric attributes, but for a regression process, we need all attributes to be numerical.

#### [5] IV. METHODOLOGY

##### POLYNOMIAL REGRESSION

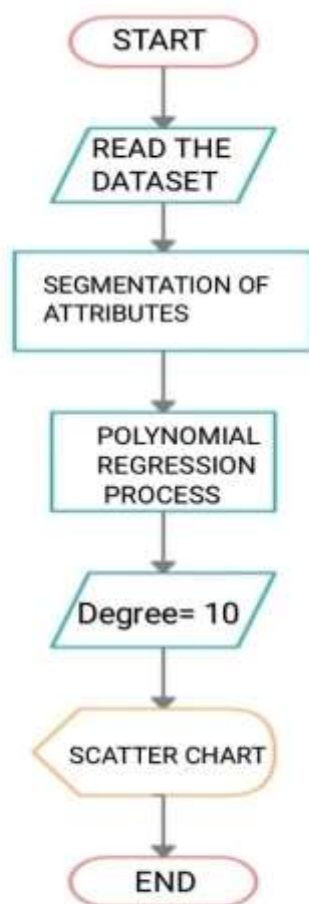
Polynomial Regression is a regression algorithm that models the relationship between a dependent (y) and independent variable (x) as nth degree polynomial. The Polynomial Regression equation is given below:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

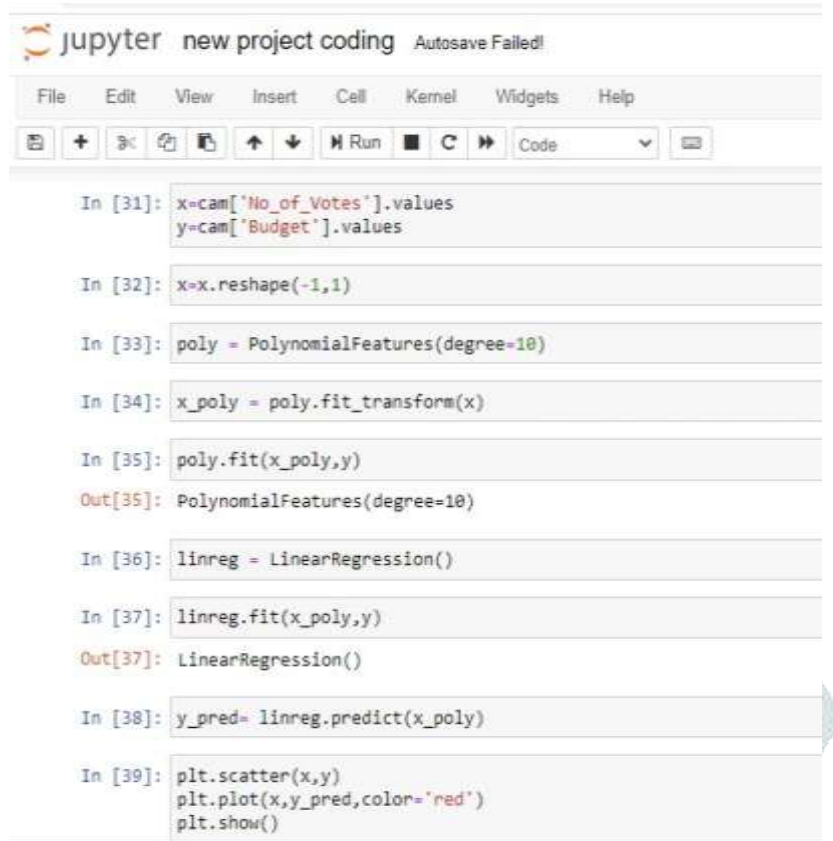
There were various regression algorithms which are generalized linear models and support vector machines [9]. It is a linear model with some modification in order to increase the accuracy. The dataset used in Polynomial regression for training is of non-linear nature [7].

Import the important libraries and the dataset we are using to perform Polynomial Regression.

##### FLOWCHART OF THE PROCESS



## [6] V.RESULT



```

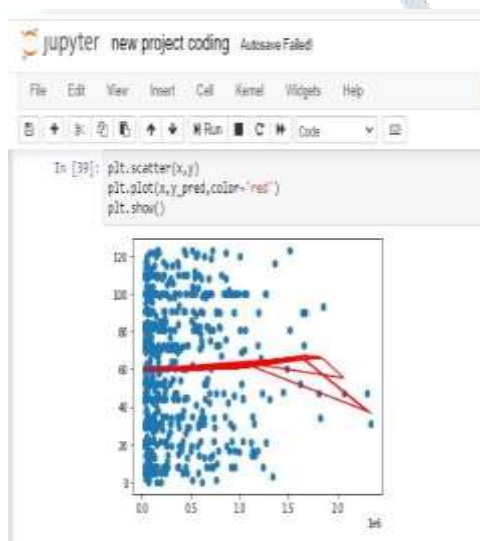
jupyter new project coding Autosave Failed
File Edit View Insert Cell Kernel Widgets Help
+ -> <-> Run C Code
In [31]: x=cam['No_of_Votes'].values
        y=cam['Budget'].values
In [32]: x=x.reshape(-1,1)
In [33]: poly = PolynomialFeatures(degree=10)
In [34]: x_poly = poly.fit_transform(x)
In [35]: poly.fit(x_poly,y)
Out[35]: PolynomialFeatures(degree=10)
In [36]: linreg = LinearRegression()
In [37]: linreg.fit(x_poly,y)
Out[37]: LinearRegression()
In [38]: y_pred= linreg.predict(x_poly)
In [39]: plt.scatter(x,y)
        plt.plot(x,y_pred,color='red')
        plt.show()

```

Fig 1

In the Fig 1, We proceed to examine the distributions of the values for two film statistics: Production budget, and total number of user votes. X is determined as No of votes and Y as Budget. PolynomialFeatures generates a new matrix with all polynomial combinations of features with given degree as 10.

Fig 2



A scatter chart is a mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. It suggests that the blue point as votes and red lines as accuracy found. As goal of plotting up of graph is to predict the success or failure of a movie by votings given by movie watchers. A statistical summary of the input variables is provided showing that values are numeric and range approximately on X axis from 0 to 2.0 and Y axis from 0-120.

## [7] VI. CONCLUSION AND FURTHER WORK

In this paper, Jupyter Notebook can illustrate the analysis process step by step by arranging the stuff like code, images, text, output etc. ...It helps a data scientist to document the thought process while developing the analysis process. One can also capture the result as the part of the notebook. One major feature of the Jupyter notebook is the ability to display plots that are the output of running code cells. However, there are other features that would be nice to have to better streamline your work and this is where the guys at Jupyter introduced Jupyter Lab. It's a full featured IDE that has everything we ever wanted to be in Jupyter notebooks.

**REFERENCES:**

- [1] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, 2nd ed. New York: Wiley, 1973
- [8] Sitaram Asur and Bernardo A. Huberman, Predicting the Future with Social Media, <http://arxiv.org/abs/1003.5699>, March 2010
- [9] Michael T. Lash and Kang Zhao, Early Predictions of Movie success, 2015
- [10] Darin Im, Minh Thao, Dang Nguyen, "Predicting Movie Success in the U.S. market," Dept. Elect. Eng, Stanford Univ., California, December, 2011
- [11] Javaria Ahmad ; Prakash Duraisamy ; Amr Yousef ; Bill Buckles "Movie Success Prediction Using Data Mining", 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), PP 1- 4. ohn Mangan and Chandra Lalwani - GLOBAL LOGISTICS AND SUPPLY CHAIN MANAGEMENT (THIRD EDITION) – Published 2016 John Wiley and Sons Ltd. – ISBN 9781119117827.
- [12] Susmita S. Magdum ; J. V. Megha "Mining Online Reviews and Tweets for Predicting Sales Performance and Success of Movies", 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) PP 334-339. Mark Barratt, Adegoke Oke - ANTECEDENTS OF SUPPLY CHAIN VISIBILITY IN RETAIL SUPPLY CHAINS: A RESOURCE-BASED THEORY PERSPECTIVE - JOURNAL OF OPERATIONS MANAGEMENT – Available online at [www.sciencedirect.com](http://www.sciencedirect.com).
- [7] ^Stevenson, Christopher. "Tutorial: Polynomial Regression in Excel". [facultystaff.richmond.edu](http://facultystaff.richmond.edu). Retrieved 22 January 2017.
- [8] Nahid Quader, Md. Osman Gani, Dipankar Chaki, and Md. Haider Ali, "A Machine Learning Approach to Predict Movie Box-Office Success," in 20th International Conference of Computer and Information Technology (ICCIT), December 2017.
- [9] Yun Wang and Sudha Ram, "Predicting Location- Based Sequential Purchasing Events by Using Spatial, Temporal, and Social Patterns", IEEE Intelligent Systems, May/June 2015.
- [10] Nanlin Jin, Peter Flach, Tom Wilcox, Royston Sellman, Joshua Thumim, and Arno Knobbe, "Subgroup Discovery in Smart Electricity Meter Data", IEEE Transactions On Industrial Informatics, Vol. 10, No. 2, May 2014.