ANALYSIS ON MOVIE OVERVIEW USING POLYNOMIAL REGRESSION

Dr.M.Rajeswari

Assistant Professor,

Department of B.Com (Business Analytics),

PSGR Krishnammal College for Women, Coimbatore, India.

rajeshwarim@psgrkcw.ac.in

Jenani.A

UG Scholar,

Department of B.Com (Business Analytics),

PSGR Krishnammal College for Women, Coimbatore, India.

janujenani909@gmail.com

ABSTRACT:

This paper is about analysis on movie overview using polynomial regression. By noticing whether the performance of particular movie is liked by all kinds of people in rural city, town. Analysis on overview the movie director can predict through the box office that the movie reached success or not. Every year more than hundreds to thousands of movies are released to the public audience with the hope that the movies getting released will be the next block buster or hit ,reached to success or failure according to the moives acceptance by the public. Statistics says that only ten to twelve movies out of twenty movies gets unprofitable, only one third of the movie gets success.

Keywords : Machine Learning, success or failure, overview, Prediction ,polynomial regression

I. INTRODUCTION

Jupyter notebooks have seen enthusiastic adoption in the data science community, to an extent where they are increasingly replacing Microsoft Word as the default authoring environment for research. Within digital humanities literature, one can find references to Jupyter notebooks (split off from *iPython*, or interactive Python, notebooks in 2014) dating to 2015.Jupyter Notebooks have also gained traction within digital humanities as a pedagogical tool. Multiple Programming Historian tutorials such as Text Mining in Python through the HTRC Feature Reader, and Extracting Illustrated Pages from Digital Libraries with Python, as well as other pedagogical materials for workshops, make reference to putting code in a Jupyter notebook or using Jupyter notebooks to guide learners while allowing them to freely remix and edit code. The notebook format is ideally suited for teaching, especially when students have different levels of technical proficiency and comfort with writing and editing code.

Movie reviews based on the viewers are an important way to gauge the performance of a movie. A movie review tells us about the the strong and weak points of the movie and deeper analysis of a movie review can tell us if the movie in general meets the expectations of the reviewer. Using analysis on film telecasting, The producers, studios, investors, sponsors in the movie industry are alike interested in predicting the box office success of the movie. we can find the state of mind of the reviewer while providing the review and understand if the person was "happy", "sad", "angry" and so on[1].

This paper work shows and analyse the film genre, the release date, the release month of movies, the languages and location from the movie dataset. The attributes like (certificate, runtime, genre, overview, IMDB ratings, Meta score, and number of votes, gross sales budget and location of a movie) taken from the dataset and the derived attributes (release month of the movie derived from release date of movie and profit from budget and revenue) is analysed to determine the movie performance.

OBJECTIVE

To Analyse on moive overview and to see whether moive reached to success or not.

- Find which moive got super hit.
- Predict and overview through box office.
- Find profit comparesion and which moive gets more profit compare to other.

II.   RELATED WORK

The main objective of our system is to recommend fims to our users based on their viewing history and ratings they provide. The System will also recommend various e-commerce companies to promote their products[3]. Depending on the genre of films of their choice to specific customers.

Personalized recommendation engines help Collaborative filtering and Content-based filtering is the predominant approach for providing recommendations to users[5]. Costin-Gabriel Chiru et al proposed Movie Recommender, a system which uses the information known about the user to provide movie recommendations[1]. This system attempts to solve the problem of unique recommendations which results from ignoring the data specific to the user.

In this objective, we try to classify whether a person liked the movie or not based on the review they give for the film. This is particularly useful in cases when the creator of a movie wants to measure its overall performance using reviews that critics and viewers are providing for the movie. Another application of this object would be to find a group of viewers with similar movie tastes (likes or dislikes). where each review is labelled with the rating of the movie on scale of 1-10. As sentiments are usually bipolar like good/bad or happy/sad or like/dislike, we categorized these ratings as either 1 (like) or 0 (dislike) based on the ratings .The film industry plays a major role in the planetary or world-wide economy.

The system also has a provision for user to select attributes on which he wants the movie to be recommended. Luis M Capos et al has analyzed two traditional recommender systems i.e. content based filtering and collaborative filtering[4].This is an efficient technique based on Hierarchical clustering for recommender system. Urszula Kużelewska et al proposed clustering as a way to deal with recommender systems[6].Two methods of computing cluster representatives were presented and evaluated.
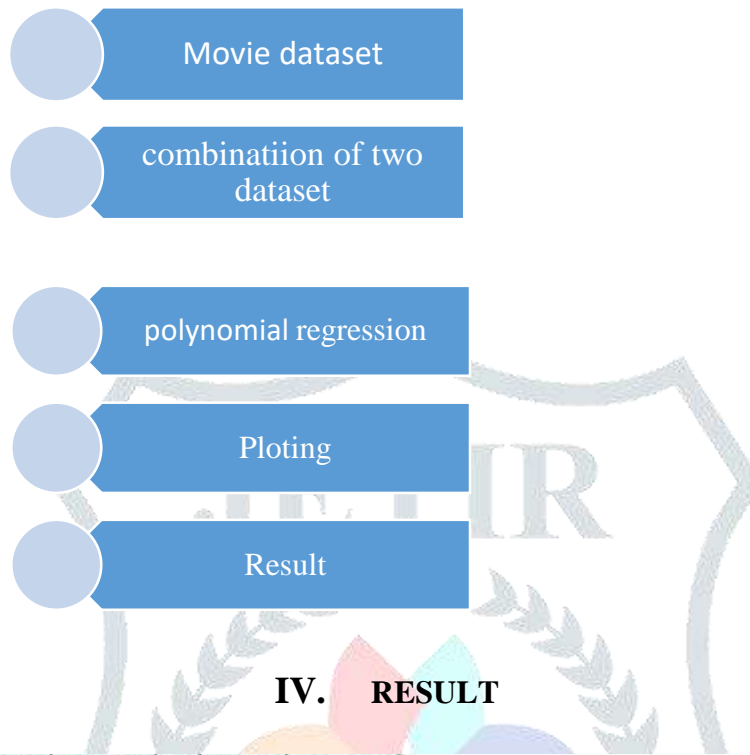
III.   METHODOLOGY

POLYNOMIAL REGRESSION:

A polynomial curve is a curve that can be parametrized by polynomial functions of R[x], so it is a special case of rational curve. Therefore, any polynomial curve is an algebraic curve of degree equal to the higher degree of the above polynomials P and Q of a proper representation. The goal of polynomial regression is to model a non-linear relationship between the independent and dependent variables (technically, between the independent variable and the conditional mean of the dependent variable).If data points clearly will not fit a linear regression (a straight line through all data points), it might be ideal for polynomial regression. Polynomial regression, like linear regression, uses the relationship between the variables x and y to find the best way to draw a line through the data points.

R-Squared: It is important to know how well the relationship between the values of the x- and y-axis is, if there are no relationship the polynomial regression can not be used to predict anything.The relationship is measured with a value called the r-squared.The r-squared value ranges from 0 to 1, where 0 means no relationship, and 1 means 100% related.Python and the Sklearn module will compute this value for you, all you have to do is feed it with the x and y arrays:

FLOWCHART OF THE PROCESS

Movie dataset

combinatiion of two dataset

polynomial regression

Ploting

Result

IV. RESULT

```
#To list the columns and column values count
for column in data.columns.values.tolist():
    print(column)
    print (data[column].value_counts())
    print("")
```

```
Flop        1
1.4         1
Name: Rating, Length: 76, dtype: int64

Verdict
Flop          343
average       187
Average        69
Hit            35
Blockbuster    25
flop           19
unknown         7
super Hit       5
super hit       4
Super Hit       2
hit             1
Name: Verdict, dtype: int64

No_of_Votes
63516       1
```

Fig: 1

In this fig 1 it show the data column value counts like rating,superhit, blockbuster,flop, average, verdict from this the moive producer can predict the performance .By giving the print column statement it displays the column.
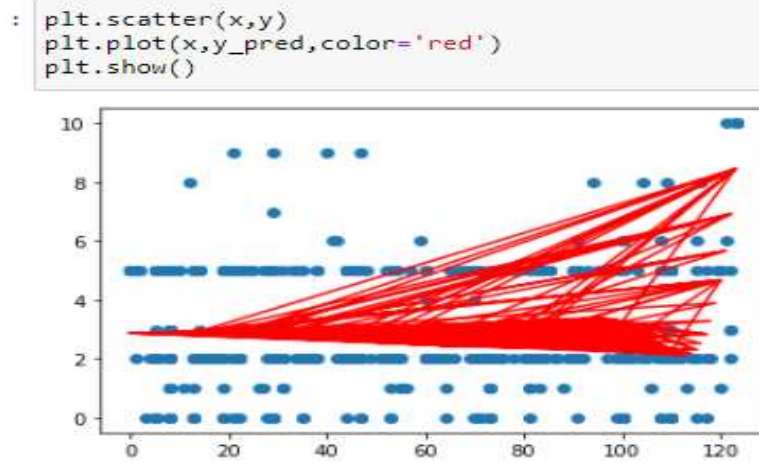
```
: plt.scatter(x,y)
  plt.plot(x,y_pred,color='red')
  plt.show()
```


Fig: 2

In this above fig 2 express the overview relationship between the movie variables x and y x variables are plotted in red colour and y variables are potted in blue in which the overview of the movie reaches ups and downs .to find the best way to draw polynomial line through the data points.

V.   CONCLUSION AND FURTHER WORK

In this paper incorporates the analysis and Prediction made on the movie dataset. This analysis helps to find the relationship between the attributes which was given in the dataset. It also helps to predict the movie result conditions with the help of python in jypter notebook. Analysis of the movie dataset shows that majority of the movies have runtime between 90 and 120 minutes. We also saw that ratings lie between 6 and 7 with mean value of 6.72. from the overview of the movie dataset by using polynomial regression the directors, producers, studios, investors, sponsors in the movie industry are alike interested in predicting the box office success of the movie.

REFERENCE:

[1] Costin-Gabriel Chiru, Vladimir-Nicolae Dinu , Ctlina Preda, Matei Macri ; "Movie Recommender System Using the User's Psychological Profile" in IEEE International Conference on ICCP, 2015.

[2] Gaurangi, Eyrun, Nan; "MovieGEN: A Movie Recommendation System", UCSB.

[3] James Bennett, Stan Lanning ; "The Netflix Prize", In KDD Cup and Workshop in conjunction with KDD,2007

[4] Luis M. de Campos, Juan M. Fernández-Luna *, Juan F. Huete, Miguel A. Rueda-Morales; "Combining content-based and collaborative recommendations: Ahybrid approach based on Bayesian networks",International Journal of Approximate Reasoning, revised 2010.

[5] Mohammad Yahya H. Al-Shamri , Kamal K. Bharadwaj; "A Compact User Model for Hybrid Movie Recommender System " in International Conference on Computational Intelligence and Multimedia Applications 2007.

[6] Urszula Kużelewska; "Clustering Algorithms in Hybrid Recommender System on MovieLens Data", Studies in Logic, Grammar and Rhetoric, 2014.

[7] Andrew L Mass, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts (2011). Learning Word Vectors for Sentiment Analysis.

[8]Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", *Advances in Artificial Intelligence*, 2009.

[9]Wu Yan, Shen Jie, Gu Tianzhu, CHEN Xiaohong, LI Hui and ZHANG Shu, "Algorithm for Sparse Problem in Collaborative Filtering", *Application Research of Computers*, vol. 24, no. 6, June 2007.

[10] Booker, E.: 'Predictive Analytics at Work at Ebay, Redbox.', in Editor: 'Book Predictive Analytics at Work at Ebay, Redbox.' (2014, edn.).