

ANALYSIS ON MOVIE RATINGS USING SUPERVISED LEARNING

Dr.M.Rajeswari¹

Assistant Professor, Department of B.Com (Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
rajeshwarim@psgrkcw.ac.in

S.Hail Evangeline²

UG Scholar, Department of B.Com (Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
haileva213@gmail.com

ABSTRACT:

Movie ratings and reviews at sites such as IMDb are commonly used by moviegoers to decide which movie to watch or buy next. Currently, viewers base their decisions as to which movie to watch by looking at the ratings of movies and also reading some of the reviews at IMDb. Moviegoers can then make a choice on which movie to observe next by selecting those movies having emotion maps with certain emotion map patterns desirable for them. Data science has the art of producing and marketing entertainment at levels never before seen. The field of knowledge science also pops up as meaty material during a sort of films that use predictive analysis, machine learning, and AI as central plot themes. Data science is involved in practically every step of the process, and in this project the analysis on ratings under various emotion contents were examined on the basis of audience acceptance towards the movie.

Keywords – Machine Learning, Emotion Analysis, Data Science, Ratings, Predictive analysis

I. INTRODUCTION

The movie to be released on that particular day may be analysed in the beforehand and certain modifications can be made in the production team in relation to budget. Movie's trailer or some songs of the film would have been released before the final release of the film. The audience ratings and their acceptance towards the movie can be easily analysed on the basis of ratings and reviews. So the needed modifications can be made in respect with the budget or production team to improve or to maintain the quality of the film.

Regression is a method of dependent values based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Machine learning may be a sort of AI that gives computers the power to find out without being explicitly programmed. The algorithm that is used for analysing the data is Linear Regression. Linear regression is a type of an analysis in which the number of independent variables is one and there is a linear relationship between the dependent and independent variables. It is commonly used for predictive analysis and modelling. This analysis helps the production team to improve or make any informed decisions regarding the film to be released.

OBJECTIVE:

- To analyse the ratings and review of the movie and to gain insight on it.
- Examine the portfolio of movie recommendations to the viewers and make any informed decisions on the same.

II. RELATED WORK

Collaborative filtering is one of the most effective and adequate techniques used in recommendation. The fundamental aim of the recommendation is to provide prediction of the different items in which a user would be interested in based on their preferences.[4] Recommendation systems based on collaborative filtering techniques are able to provide approximately accurate prediction when there is enough data.[2]

User based collaborative filtering techniques have been very powerful and successful in the past to recommend the items based on the user's preferences. But, there are also some certain challenges such as scalability and sparsity of data which increases as the number of users and items increases. In a large website, it is difficult to find the interested information in a certain time. But the recommendation system filters out information and items that are best suitable for us. Although there are different recommendation approaches, yet collaborative filtering technique is very popular because of the effectiveness.[1]

For item-based recommendation in the paper, they have used the ratings of those movies that are highly similar to the rating of the movie, which is provided by a proper user from using the item similarity weights in the item similarity weight matrix. [5].

In 1980, J.A. Russell characterized emotions in two dimensions, namely, activation and pleasure, and proposed the circumplex model [6]. With his model, Stone provided a database containing lexicon of emotions, with more than 100 categories and 11,000 words [7]. Recently, Cambria et al proposed the hourglass model [3] which is a more advanced model than the circumplex model, and has four independent, but concomitant, dimensions, namely, pleasantness, attention, sensitivity, and aptitude.

III. METHODOLOGY

Linear regression is one of the most important and widely used regression techniques which helps us to find the relationship between the independent and dependent variables.. It is among the simplest regression methods. One of its advantages is the ease of interpreting results. The numpy package and matplotlib are imported to plot the graphs. The linear regression data, IMDB Rating is fetched and given its datatype as an object. The X and Y axis of array values were ordained and to find the linear regression value, initiating -1 and 1 to predict the output. Using the scatter plot to visualize the data, IMDB ratings , Year and Type of emotion or Genre have been analyzed.

Ratings are designed to say what kind of content is included in the movie and not designed to indicate that films are appropriate or inappropriate to give parents a chance to take decisions on what they are allowing their children to watch. Ratings are assigned by parents who consider various factors such as violence, sex, language and drug use, then assign a rating.

Here, we have analysed the ratings of the movie and also by the year. The changes occur gradually year by year on different perspectives of the audiences on the basis of Type of the movie (Tragedy, Drama, Commercial, Comedy, History, Horror, and Love), Genre, The Direction and Cast and so on. A set of audience would like a different genre and type and they may be eager to watch the film. So when a trailer or a song is released before the final release of the film a majority of the people will rate the movie accordingly. The production team may be able to make any informed decisions on the beforehand. By means of the following analysis we might get a clear insight on all these factors in the movie.

There are five basic steps for implementing linear regression:

STEP -1 Import the packages and csv file you need.

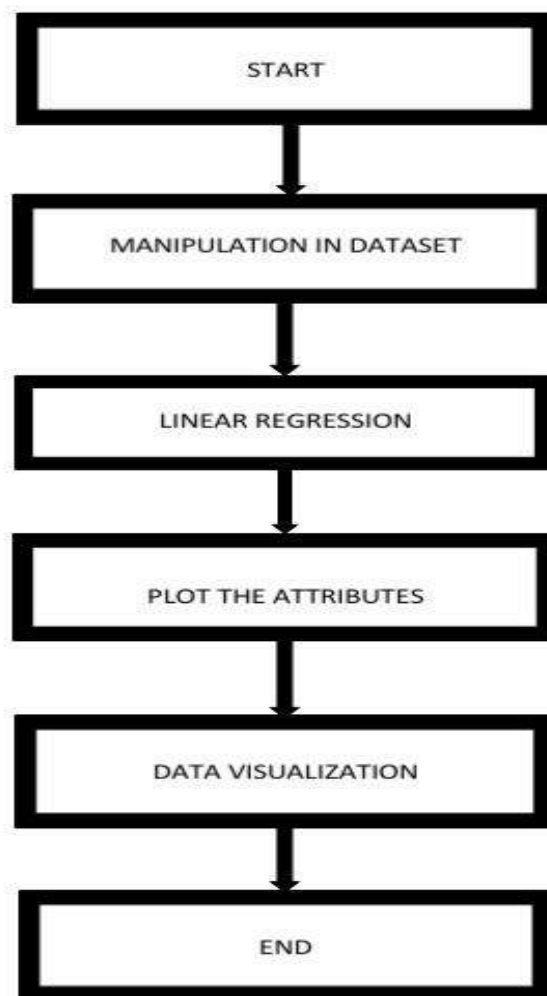
STEP-2 Provide data to work with and eventually do appropriate transformations.

STEP -3 Create a linear regression model and fit it with existing data.

STEP-4 Check the results of the model to know whether the model is satisfactory.

STEP -5 Apply the model for predictions. [8]

FLOWCHART OF THE PROCESS



IV. RESULT

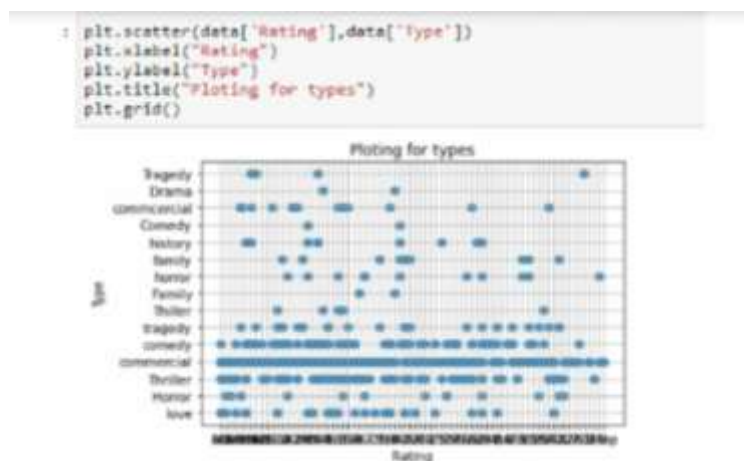


Fig 1

In the Fig 1, The graph has been plotted for Types of movie emotions and the Ratings which the viewers rate accordingly. It is clearly evident that among all other emotions and type Commercial films have got the highest ratings and are mostly accepted by the audience and viewers.

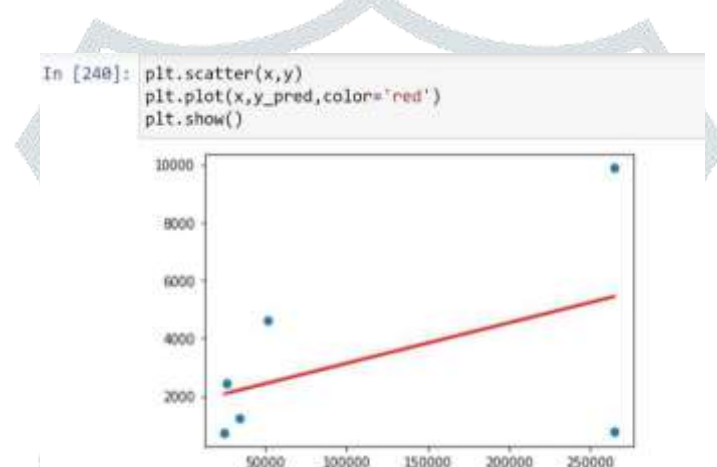


Fig 2

In the above Fig 2, A scatter chart with the regression model is an excellent tool which can be used to depict the relationship between two variables. This shows the relationship between two variables visually. When the Ratings of the movie goes upward, a high number of audience likes it and these kinds of movies will definitely be successful in the upcoming days.

FINDINGS:

- It may be useful for users in their decision-making process to choose the next movie to watch if a movie also comes with an (expected) emotion signature or an emotion map type.
- Clearly, once emotion maps for all movies are at hand, if a user perhaps submits his desired emotion state and, possibly, the desired genre of the movie, it is easy to build a personalized movie recommender system for each user.
- It is useful for an initial starting point to a customized movie portfolio for a user. If a user writes a review and rates a movie with a good score; we calculate his emotions; then our emotion maps can recommend to him other movies based on his past ranking/emotion trend using other similar user reviews.

V. CONCLUSION AND FURTHER WORK

In this paper, Jupyter Notebook is used as a tool for capturing the emotional content of the movies via the emotions expressed in the films. Linear regression, a simple statistical supervised learning algorithm has made it even more easier to analyse the relationship between the basic emotions or the movie type and its rating by the viewers. Different clusters of emotions are aggregated and the majority of the acceptance towards the movie will definitely lead to a good success of the film.

REFERENCES:

- [1] Acker, Ally (1991). Reel Women: Pioneers of the Cinema, 1896 to the Present. New York: Continuum. ISBN 0-8264-0499-5
- [2] Billsus, D., and Pazzani, M. J.: 'Learning Collaborative Information Filters'. Proc. Proceedings of the Fifteenth International Conference on Machine Learning 1998.
- [3] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7403 LNCS, 2012, pp. 144–157.

- [4] Group, C.C.: 'Distribution of movie and TV rental market revenue in the United States from 2012 to 2016', in Editor: 'Book Distribution of movie and TV rental market revenue in the United States from 2012 to 2016' (2017, edn.).
- [5] Kharita, M. K., Kumar, A., & Singh, P. (2018). Item-Based Collaborative Filtering in Movie Recommendation in Real-time. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). DOI:10.1109/icsccc.2018.8703362.
- [6] J. A. Russell, "A circumplex model of affect." Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980. [Online]. Available: <http://psycnet.apa.org/journals/psp/39/6/1161>
- [7] P. Stone, "General Inquirer." [Online]. Available: <http://www.wjh.harvard.edu/inquirer/Home.html>
- [8] Mirko Stojiljković - LINEAR REGRESSION IN PYTHON - IMPLEMENTING LINEAR REGRESSION IN PYTHON - <https://realpython.com/linear-regression-in-python/>

