TV SHOWS POPULARITY USING DATA MINING

[1]Hasan Parvez Siddiquee, [2]Sandeep Yadav, [3]Poonam Jain

[1]Post Graduate Student, [2]Post Graduate Student, [3]Assistant Professor

[1] Department of Information Technology,

[1]Thakur College of Science and Commerce, Mumbai, India

Abstract: Rapid advancement in technologies, easy accessibility of resources and high demand of TV shows in the world of entertainment. Led to large number of productions of TV shows. Each one has built their own regime in the world of entertainment. Nowadays number of productions and cost of production is very high due to which it become most important and challenging tasks to predict TV Show Popularity using some simple method over traditional one which is costly and time consuming. This study has been undertaken to perform detail analysis of ongoing tweets flowing on twitter related to different TV Shows on Indian television. Using python scripts, we can tell you the sentiments of people regarding to any TV Shows by analyzing tweets related to that shows. Proposed python script uses many in-build libraries like Tweepy, Text Blob and Matplotlib etc. these packages helps us to implement Twitter's API any many other functionality to search for tweets about any TV Shows and analyze each tweet to see how positive or negative its emotion is**.** Based on sentiment about different shows we determine the popularities of shows among viewers.

Index Terms **- Tweets, Popularity, Analysis, TV Shows.**

I. INTRODUCTION

With the rapid development of sharing Websites, more and more people would like to become audiences in their daily entertainments. such as TV shows and Web sites, to attract more audiences. although many efforts have been taken for the popularity prediction. Also, episodes released on weekends or holidays may attract more audiences than those on workdays. Additionally, various episodes are released on distinct days. Thus, the prediction of popularity is an important task. A simplest means of prediction is based on people's rating and choices and comments. Easy importing of data and exporting it into the graph. Graphical data in the printable format. The visitor will get to know the show popularity. Reality TV is the mantra for the next generation of TV producers and executives. The main purpose of this is to find out TRP ratings. Most reality shows these days are concerned in dance, singing and acting related shows. We conclude to build such a system that will recognize people's sentimental comments on TV shows. The tweets related to the particular show will be extracted. The comments will be gathered from various sources social networking websites like Twitter. On the basis of people's comment and the TV Show popularity will be rated accordingly.

II. LITERATURE REVIEW

D. Anand, A.V.Satyavani, B.Raveena and M.Poojitha implemented predict the TV show popularity rating. They use two algorithms K-Means and incremental K-Means for analysis of TV show popularity rating. After performing classification and clustering. they have found their best results are achieved through Incremental clustering algorithm at 97% accuracy. Which is very much acceptable. This implementation can easily calculate TRP rating with the help of these clustering techniques despite the traditional methods and can gain more efficiency and time [1]. Tejaswi Kadam, Gaurav Saraf, Vikas Dewadkar, and P.J Chate had fetch tweets from social media website like twitter for sentiment analysis. In their research they had use NLP (natural language processing) deep learning models to make computer understand the user's tweet in their fancy texting style. Porter Stemmer algorithm is also used to construct the tweet in well formed English language to understand user's sentiments [2].

III. STUDY AREA AND CHARACTERSTICS

As the demand of television entertainment is increasing, several TV Show of different genre releases in India every year. Top TV channels in India are known for their prime-time (6-11 am) shows. We have covered prime-time shows from top rated channel in India. Like Star Plus, Sony TV, Sony Sab and Colors etc. Tweet related to numbers of prime time shows from top rated channels are retrieved from twitter's tweet using twitter API. These data are being stored in excel sheet for in-depth analysis.

IV. METHODOLOGY

The methodology is the actual implementation of propose system where the idea is being implemented and tested on the data and the performance and result of implementation is suggested. So, here we are using Python script to implement our proposed model. Python is very powerful and well reputed language for data science. several libraries are contributed by python community to design well structured solution for any problem in the world of data science. The complete exposure of proposed solution is implemented by the following adopted methodologies which are as follows:

A. RAW DATA COLLECTION

To perform detail analysis of popularity of TV Shows we need a large amount of relevant data. Traditional but best way to acquire such amount of data is fetching viewer's tweets about TV Shows from social media. Here., We have created a twitter API and written a python script using well known library called tweepy. This library let us to fetch all the on-flowing tweets on twitter using keywords associated with particular TV Shows. These tweets will be stored in excel sheets for analysis.

B. SENTIMENT ANAYSIS

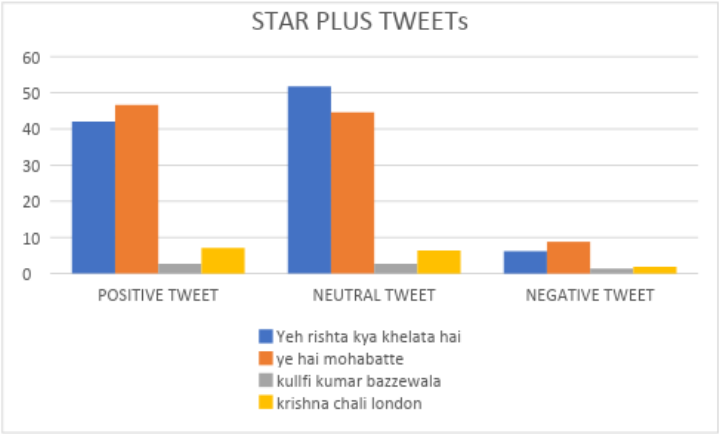
After acquiring and storing relevant data in excel sheet, we are importing a library called Text Blob in our script to measure the sentiments of each and every tweet fetched using keywords for particular TV Shows. Text Blob is very powerful and well-defined library for processing textual data. A common API is provided for diving into common into common natural language processing (NLP) tasks such as speech tagging, translation, noun phrase extraction, classification, sentiment analysis, etc. There are some parameters that we have define as positive, negative and neutral for classification of tweets sentiment. Script will determine the sentiment of tweets and classify it in respective classes.

C. PLOTING THE SENTIMENT

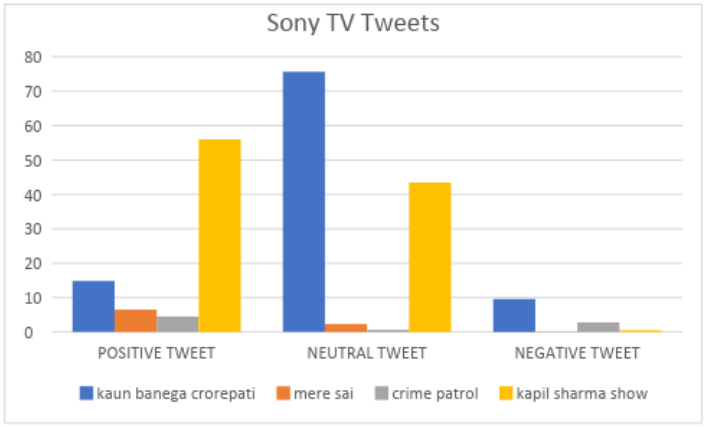
As we have sense the sentiment of tweets using Text Blob we need to visualize the resultant data in graphical representation. As python is well reliable and reputed language for data science .it provide many libraries for data visualization. Here, we have used matplotlib to plot the result on 2D plane. The graphical representation of data classifies sentiment of tweets in three main class i.e., positive, negative and neutral.

D. POPULAR TV SHOWS

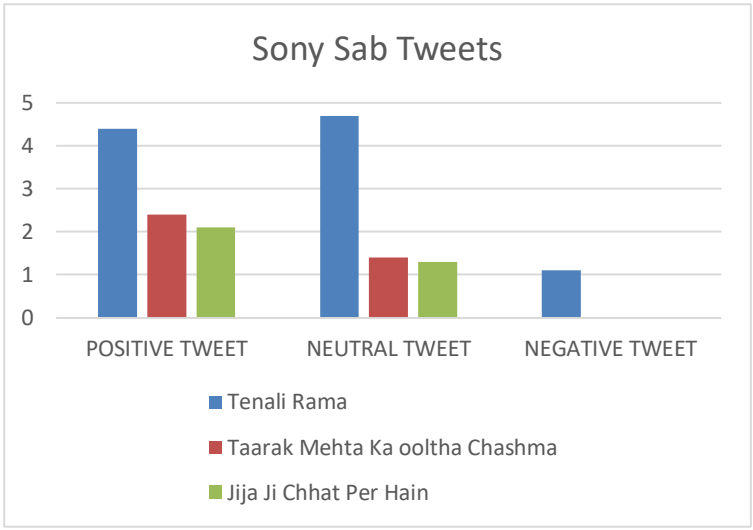
The pie-chart in the above step is displaying the sentiment regarding TV Show. to compare the popularity among different shows we have created an excel sheets where we have mention the tv shows and respective tweets sentiment in the form of positive negative and neutral.
Here is the screenshot of graph generated from sample dataset.



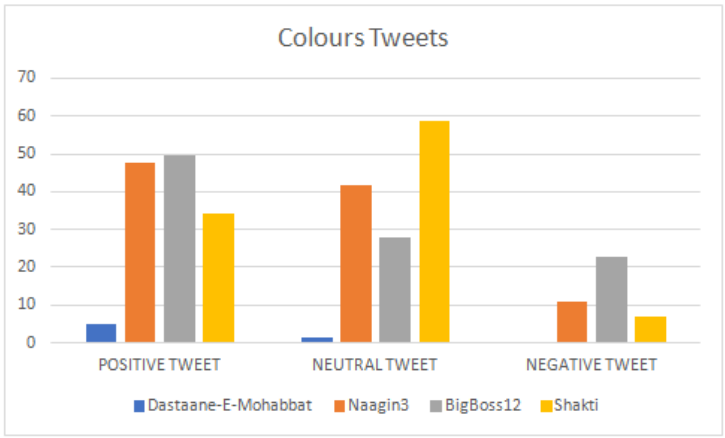
According to the above table we can assert that the TV shows call "Yeh Rishta Kya Kehlata" is the most popular one among all the show running on channel Star Plus. Running back "Yeh Hai Mohabbatein" is also consider as well design competitior of Yeh Rishta Kya Kehlata. There is high probability that "Yeh Hai Mohabbatein" will left "Yeh Rishta Kya Kehlata" back in future.



There many TV shows run on channel called Sony TV. Each show has their own interesting content to target potential viewers. From the plotted graph we can say that most of the people are talking positively about the "The Kapil Sharma Show". Most people have strong neutral sentiment regarding "Kon Banega Crorepati". So, both the shows are going hand on hand. It is very difficult to predict the popularity between these two shows Currently running on Sony TV.



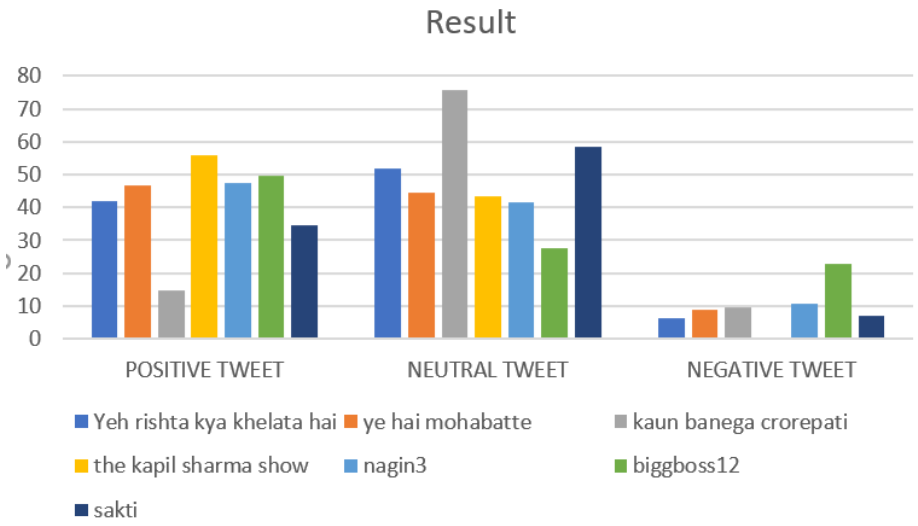
Talking about Sony Sab, It is well known channel in Indian Television because of its longest running shows called "Tarak Mehta Ka Ulta Chasma". But recently released show called "Tenali Rama" is the leading over "Tarak Mehta Ka Olta Chasma". Most of the people shows more positive sentiments has been shown toward "Tenali Rama" as compared to "Tarak Mehta Ka Olta Chasma". Neutral comments on "Tarak Mehta Ka Olta Chasma" is much more than any other running TV Shows on Sony Sab. Coming to "Jija Ji Chhat Per Hain" It is a 3rd Running show on Sony sab after "Tenali Rama" and "Tarak Mehta Ka Olta Chasma"



As the "Shakti" received the more neutral comment as compared to any other shows flowing on Colour and positive comment approx. 34%. it stood 1st in the popularity on this channel. However, Nagin is doing well and hold same level of popularity. It has shaded great impression on the viewers of Colour Television. When we talk about Big boss. it has received the more positive comment as compared to any other shows on this channel. But it is topping in pouring bad impression on TV viewers. negative comments approx. 22% clearly shows the Big boss is not that effective as other shows are.

V. RESULT

Here we have selected only top-rated TV Shows running on Indian Television. After detail analysis of classified tweets, we can conclude that the "The Kapil Sharma Show" the most popular tv show dominating the entertainment industry. Followed by the "The Kapil Sharma Show" "Yeh Rishta Kya Kehlata Hai" holds second place in terms of popularity among Indian Television shows. "Shakti" and "Yeh Hai Mohabbatein" both are receiving same level of positive and neutral tweets i.e., these both are having same level of popularity.



VI. CONCLUSION

We have collected data from Twitter using tweepy. Once the data is extracted and loaded using Python script, we have analyzed the views and sentiments of the people. Based on these sentiments, we are now successfully able to classify the tweets as positive tweet, negative tweet or neutral tweet. These classified tweets help us to analyze the popularity of different TV Shows Currently Running in India

FUTURE ENHANCEMENTS

As part of future enhancements, we can predict the TV show popularity rating. We can easily calculate TRP rating with the help of clustering techniques which can gain more efficiency and time. After performing classification and clustering, we can find the best results which can be achieved through Incremental clustering algorithm. The attributes can then be contributed to most of the information like number of shows, view count for each channel, etc. The channel and the show that would get the highest TRP rating will then be awarded as the most viewed show all over. Else by using the training methods, calculating TRP would be time consuming, costlier and technically and economically infeasible.

ACKNOWLEDGEMENT

We take this opportunity to express our gratitude to Thakur College of Science and Commerce for giving us this golden opportunity to accomplish this project.

We are very thankful to the IT Department and Dr. Santosh Kumar Singh, Head of Department of Information Technology for his kind co-operation in the completion of this project.

We would like to express special vote of thanks to our project guide, Mrs. Poonam Jain who gave us this opportunity to do this wonderful project and has also helped us a lot in the course of this research and we came to know about so many new things. We are thankful and indebted to her.

REFERENCE

[1] Kadam, Tejaswi, et al. "TV Show Popularity Prediction using Sentiment Analysis in Social Network." *International Research Journal of Engineering and Technology* 4.11 (2017).

[2] D. Anand, A.V.Satyavani, B.Raveena and M.Poojitha "ANALYSIS AND PREDICTION OF TELEVISION SHOW POPULARITY RATING USING INCREMENTAL K-MEANS ALGORITHM." International Journal of Mechanical Engineering and Technology (IJMET).

[3] Ali, Kamal, and Wijnand Van Stam. "TiVo: making show recommendations using a distributed collaborative filtering architecture." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.

[4] Mei, Qiaozhu, Dengyong Zhou, and Kenneth Church. "Query suggestion using hitting time." *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008.

[5] Ratkiewicz, Jacob, et al. "Characterizing and modeling the dynamics of online popularity." *Physical review letters* 105.15 (2010): 158701.

[6] Xin, Yu, and Harald Steck. "Multi-value probabilistic matrix factorization for IP-TV recommendations." *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 2011.

[7] Cannon, Mark E. "Manipulating and analyzing data using a computer system having a database mining engine resides in memory." U.S. Patent No. 6,029,176. 22 Feb. 2000.

[8] Fernandes, Kelwin, Pedro Vinagre, and Paulo Cortez. "A proactive intelligent decision support system for predicting the popularity of online news." *Portuguese Conference on Artificial Intelligence*. Springer, Cham, 2015.

[9] Joachims, Thorsten. "Optimizing search engines using clickthrough data." *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2002.

[10] Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." *Computer*8 (2009): 30-37.

[11] Zhang, Wei, et al. "Scalable visual instance mining with threads of features." *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014.

[12] Zhang, Ya, Weiyuan Chen, and Zibin Yin. "Collaborative filtering with social regularization for TV program recommendation." *Knowledge-Based Systems* 54 (2013): 310-317.