

TV Shows Analysis using Data Mining

Siddharth Singh¹, Swapnil Gaware², Gali Nikhita³

¹⁻³Department of Computer Engineering BV (DU) COE, Pune.

ABSTRACT: Large number of productions of TV shows today is due to rapid advancement in technologies, easy accessibility of resources and high demand of TV shows in the world of entertainment. Own regime in the world of entertainment has been built by each one in the market. Traditional one which is costly and time consuming is not used as number of productions and cost of production is very high nowadays which makes it important and challenging tasks to use simple methods to predict TV Show Popularity. Analysis of ongoing tweets flowing on twitter related to different TV Shows on Indian television has been studied in this particular study of us. Sentiments of people regarding to any TV Shows by analyzing tweets related to that shows can be explained by python scripts. using many in-built libraries like Tweedy, 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 by using these python scripts. We determine the popularities of shows among viewers on different shows based in the sentiment.

INTRODUCTION

We all know how television has if not taken control or at least affecting our lives. There are ample amount of TV shows channels having plenty of shows majorly based on comedy, singing and reality-based events. We can also see that viewers who missed the episode can watch them at the end of the week; it also tells us that people are more interested in television program. Adding to that today's generation seems to be more interested in TV shows as there are shows curated for different individuals. TRP which is stands for television rating point can be found in many different ways. It is tool by which we can trace which programs are viewed most. With the help of TRP we can find the index choice of people furthermore the popularity of a particular channel. This is computed with the help of a device that is attached to TV of the no of users we want to know. These number are considered as a prototype for the overall TV shows in distinct fields. IT records the time as well as the program that one watches on an every day. They take average of 30 days which tells the status of views of channel. TRP is then compared among different TV programs. We have used k- mean and incremental k-mean algorithm to compare the TRP furthermore with the rapid sharing of Websites, more and more people would like to become part of audiences in their daily entertainments. Even though many have tried different ways for the popularity prediction. Also, episodes released on weekends or holidays attracts more audiences than those on workdays because of suspense. Additionally, different episodes are released on different days. Hence, the prediction of popularity is an essential task. It's quite

easy to import data and exporting it into the graph. Graphical data is in the printable format. The main idea here is to find out TRP ratings. Most reality shows these days are concerned in dance, singing and acting related shows. We are trying to build such a system that will recognize people's sentimental comments on TV shows any given day. The tweets referring to any particular show will also be used. The comments will be gathered from various sources social networking websites like Twitter and Facebook. On this basis of people's comment we will rate the TV shows popularity.

STUDY AREA AND CHARACTERSTICS

In India, as demand of TV entertainment increases it led to release of different genres TV shows. Prime time shows describes the most popular TV channels. Star plus, Sony TV, Sony sab and Colors are one of the top rated channel and are covered under prime time. Information of top rated channel is gathered from twitter tweet using twitter API.

Excel sheets helps in storing the data retrieved from twitter tweets.

METHODOLOGY

Methodology describes the idea of implementing and testing on the data. Python script is used for implementation. Python is very powerful and well reputed language for data science. Python community contributes several libraries to design well-structured solution for any

problem. Following methodologies shows the proposed solution:

RAW DATA COLLECTION

A large amount of data is required for performing analysis of TV shows. Tweedy is a twitter apian is created using a python script. By using keywords of particular TV shows we can fetch the tweets using the tweedy. The tweets which are fetched are stored in excel sheets for analysis.

SENTIMENT ANAYSIS

A library called Text Blob is used for recording the sentiments of each and every tweets.

Text Blob is a well-defined library for processing textual data. For achieving natural language processing (NLP) tasks such as translation, classification, sentiment analysis, noun phrase extraction, etc. an API is provided which divides it. Three parameters are classified according to tweets sentiment such as positive, negative and neutral. According to the sentiment of tweets, script are classified in particular classes.

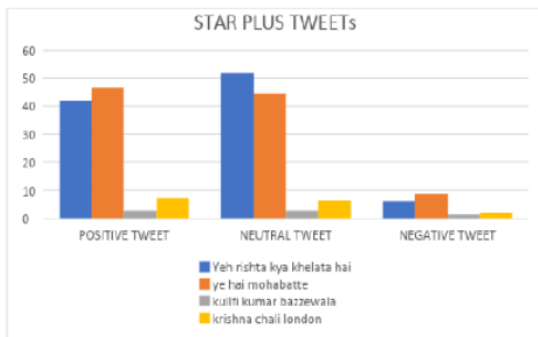
PLOTTING THE SENTIMENT

Graphical representation is used for visualization of resultant data. Python provides many libraries which can be used for data visualization. Matplotlib is used to plot the result on 2D plane. The sentiments of tweets are classified as positive, negative and neutral according to graphical representation.

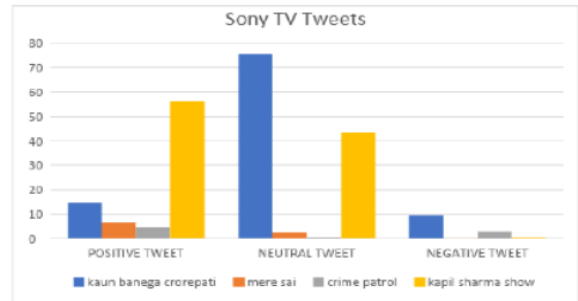
POPULAR TV SHOWS

The below pie-chart is displaying the sentiment regarding TV Show. We have created the excel sheets for comparing the popularity of TV shows.

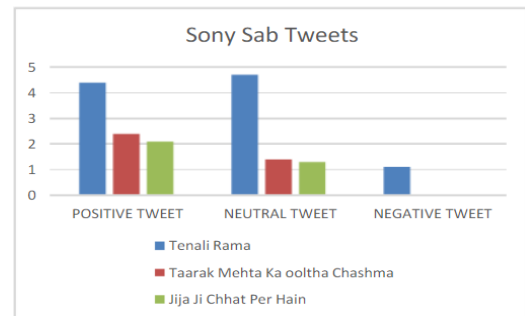
According to sentiments of tweets, we have mention the TV shows as positive, negative and neutral.



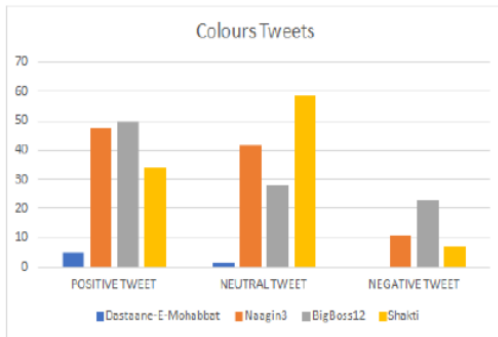
“Yeh rishta kya kehlata hai” is the most popular one among all the show running on channel star plus. Competitor of yeh rishta kya kehlata hai is yeh hai mohabbatein and it can led back to it in future.



Multiple TV shows are seen in Sony TV. From the plotted graph we can say “The kapil sharma” show is the most trending show among the audience. Positive sentiments are shown towards “Kaun banega crorepati”. It gets difficult on predicting who is most popular among these two shows on Sony TV.



One of the well-known channel in Indian TV is Sony Sab because it has longest running show called “Tarak mehta ka oolta chashma”. Recently, “tenali rama” is leading one and it has taken over “taarak mehta ka oolta chashma”. Positive sentiments are shown towards “tenali rama” as compared to “tarak mehta ka oolta chashma”. “Tarak mehta ka oolta chashma” has more neutral comments. Third most running show is “Jija ji chat par hain” after “tenali rama” and “taarak mehta ka oolta chashma”.



Shakti which is the trending show on Color has more neutral comment as compared to any other show on color. "Nagin" is the competitor show to "Shakti" on color channel. Coming to "Big Boss" it has received the more positive comment as compared to any other shows on this channel.

CONCLUSION

Most popular TV shows on Indian TV are selected according to rating. After deep analysis "The kapil sharma show" is the most popular TV show dominating the entertainment industry. "Yeh rishta kya kehlata hai" is second highest viewed TV show.

REFERENCES

[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.