

# REALITY SHOW ANALYTICS FOR TRP RATINGS BASED ON VIEWER'S OPINION

Sunil Bhutada<sup>1</sup>, Aishwarya Lakshmi<sup>2</sup>, V.Mary Shravya<sup>3</sup>, J.Sneha Reddy<sup>4</sup>

<sup>1</sup>Professor, Dept. of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

<sup>2,3,4</sup>U.G. Student (B.E), Department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

\*\*\*

**Abstract-** TV show Popularity forecast utilizing supposition investigation is a standout amongst the most fascinating and testing undertakings. Unscripted television is the new mantra of TV makers and station executives. The view of unscripted television appears from reasonability in silver screen. These shows normally collect rivalry and give cash as prizes, these shows will assemble watcher's online postings, mobile messages and break down to expand TRP appraisals. A large portion of the TV programs which are being broadcasted now a days are reality indicates gaining practical experience in dancing, singing and acting. Everything now is competition. Some indicates are acquired legitimately from abroad.

We finish up to assemble such a framework, to the point that recognize people groups wistful remarks on programs. Remarks will be extricated from twitter and will be put in CSV records. These are additionally characterized into various classes in view of their conclusion on specific TV appears. In view of their postings and conclusions, the TV Show prevalence will be evaluated in like manner. In our undertaking we break down recent TV reality show of JABARDASTH and COMEDY NIGHTS WITH KAPIL their prominence for instance.

**KEYWORDS:** Twitter API, Sentimental analysis, tweets, TV shows, Youtube.

## 1. INTRODUCTION:

Presently a days reality demonstrates have turned out to be exceptionally prominent and individuals are especially keen on watching the shows and remarking on it by utilizing different long range informal communication destinations. By breaking down people groups votes on portable messages and online postings the champ is given the value cash and in view of the people groups conclusions the prevalent show is anticipated. A large portion of the TV programs which are being broadcasted now a days are reality indicates spend significant time in dancing, singing and acting. The Indian reality demonstrates have likewise been reliably effective in

offering a wide variety, from ability chase shows to move dramas, acting flicks, talk shows, chat shows, cookery indicates etc. "The reality chase list is interminable".

Television programs and episodes draw in more audience. Episodes discharged on ends of the week or occasions may pull in more group of onlookers than those on workdays. Distinctive scenes are discharged on various days, in this way the prevalence expectation for TV demonstrates turns into a testing assignment. This is the place where sentimental analysis comes into utilization. Feeling investigation is the programmed extraction of assessments, feelings, and notions from writings. Feelings and assessments are impressions and not actualities, which are goal or unbiased. Through conclusion examination, a given content can be grouped into three classes - positive, negative, or impartial.

The fundamental motivation behind this work is to assess the execution of TV appear and furthermore figure what number of individuals enjoyed the specific show and we anticipate the ubiquity of that show in light of the tweets. we extract tweets from social media site like Twitter and Youtube.

## 2. LITERATURE SURVEY:

Tejaswi Kadam et al. <sup>1</sup> stated in their paper that, "The purpose of this work is to evaluate the performance of TV Show and also calculate how many people are liked to a particular show or actors of that show and predicting Popularity of that shows, based on the text reviews. We are getting reviews on social networking websites like Twitter."

Saraee et al. <sup>2</sup> stated in their paper that, "We gather a series of interesting facts and relationships using a variety of data mining techniques. In particular, we concentrate on attributes relevant to the user ratings of movies, such as discovering if big-budget films are more popular than their low budget counterparts, if any relationship between

movies produced during the "golden age" (i.e. Citizen Kane, It's A Wonderful Life, etc.) can be proved, and whether any particular actors or actresses are likely to help a movie to succeed. The paper also reports on the techniques used, giving their implementation and usefulness."

A.V.Satyavani et al. <sup>3</sup>stated in their paper that, "Based on the number of views we will rate that the particular Television shows accordingly with the highest Rating. The TRP can be compared among different shows and be viewed in bar graphs, pie charts, histograms. We have K-Means and Incremental K-Means algorithms to compare the TRP."

Priyank Jain. <sup>4</sup>stated in their paper that, "In this paper we present a system that uses a novel approach to analyze the audience of a TV show using machine learning, social media and big data technologies. This system will allow TV show producers and potential advertisers to perform sentiment analysis, gender analysis, device analysis of a TV show."

Ashik Abdullah. stated in his paper that, "Sentiment analysis is being used on many fronts to extract public sentiment. It can collect data automatically from microblogging sites such as twitter. This user generated data can be used for various applications. For a TV show to gain market and to quantify its success, public opinion can be extracted to find the popularity of a particular TV show. People nowadays are writing on microblogging sites about various TV shows they are watching."

### 3. PROPOSED ARCHITECTURE:

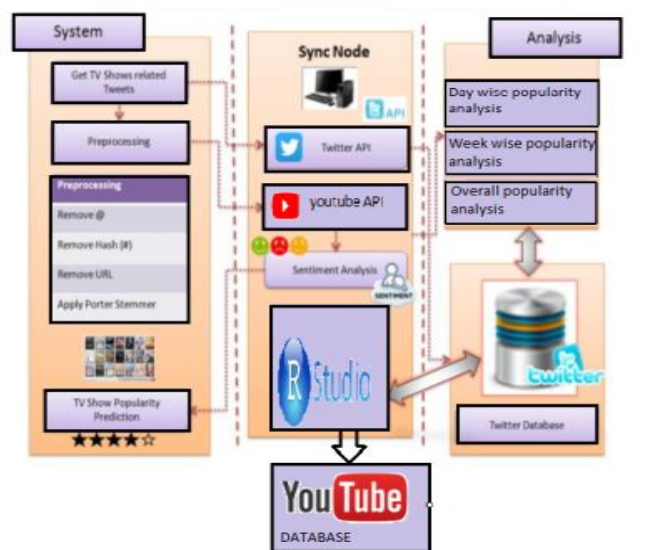


Fig -1: Architecture

The fig-1 demonstrates the engineering of the proposed framework. The framework stream is isolated into followings way. Right off the bat the framework gets TV demonstrate related information from twitter and youtube API where content audits identified with the specific show are given.

#### 3.1 Hash labeled informational index:

To make the hashtagged dataset, we first sift through copy tweets, non-English tweets, and tweets that don't contain hashtags. we examine the appropriation of hashtags and recognize what we expectation will be sets of regular hashtags that are demonstrative of positive, negative and unbiased messages. These hashtags are used to choose the tweets

#### 3.2 Pre-processing:

Pre-processing is one of the critical strides in content mining, Common Language Processing (NLP) and data recovery (IR). which gives tokenization, standardization i.e. expel @,remove #and URL. Information pre-processing is utilized to remove intriguing and non-minor information from unstructured content information. Data Retrieval is essential for choosing which records in an accumulation ought to be recovered with the goal that we can fulfill a client's requirement for data.

#### 3.3 Normalization:

To do handling on regular words, it is fundamental to perform standardization that for the most part includes wiping out the accentuation, changing over the whole content into lowercase or capitalized, changing over numbers into words, growing truncations, canonicalization of content, expels prevent words from input content information. Stop words are the word that is consequently precluded from a PC produced concordance or record.

#### 3.4 Sentimental analysis:

Sentimental analysis is a kind of information mining that measures the slant of individuals' conclusions through natural language processing (NLP), computational phonetics and content examination, which are utilized to extricate and break down subjective data from the Web - for the most part online networking and comparable sources. It is the extraction of assessments, feelings, and notions from writings. Feelings and assessments are impressions and not actualities, which are goal or unbiased. Through conclusion examination, a given

content can be grouped into three classes - positive, negative, or impartial.

### 3.5 R Studio:

RStudio is a free and open-source incorporated advancement condition (IDE) for R, a programming dialect for factual figuring and designs. RStudio was established by JJ Allaire, maker of the programming dialect ColdFusion. Hadley Wickham is the Chief Scientist at RStudio. RStudio is accessible in two releases: RStudio Desktop, where the program is run locally as a customary work area application; and RStudio Server, which permits getting to RStudio utilizing a web program while it is running on a remote Linux server. Prepackaged appropriations of RStudio Desktop are accessible for Windows, macOS, and Linux.

### 3.6 Twitter API:

Twitter is an online news and long range informal communication benefit where clients post and interface with messages, called "tweets." These messages were initially confined to 140 characters, however on November 7, 2017, the farthest point was multiplied to 280 characters for all dialects aside from Japanese, Korean and Chinese. Registered clients can post tweets, yet the individuals who are unregistered can just read them. Clients get to Twitter through its site interface, SMS or a cell phone app. Twitter, Inc. is situated in San Francisco, California, United States, and has in excess of 25 workplaces around the globe.

### 3.7 YouTube Analytics API:

The YouTube Analytics API empowers you to produce custom reports containing YouTube Analytics information. The API underpins reports for channels and for content proprietors. The Analytics API gives sifting and arranging parameters, so the calling application does not have to locally bolster those capacities. The API likewise enables you to recover information for YouTube Analytics gatherings, where a gathering is a custom accumulation of up to 200 recordings, playlists, channels, or resources.

### 3.8 Analysis :

**Day Wise Popularity Analysis:** We can break down one day fame like live show is a one day indicate so client need to discover specific date savvy investigation.

**Week savvy Popularity Analysis:** In three seasons which sort of shows was to hit in given specific period this

compose examination done in season shrewd investigation.

**Month Popularity Analysis:** We can likewise compute month shrewd famous shows rundown and watch it. Like day and week examination we break down a overall shrewd.

## 4. PROBLEM STATEMENT:

Based on the massive show information, it would be interesting to understand what are the important factors that make a show more successful than others. So, we would like to analyze what kind of reality shows are more successful, in other words, get the popularity of the show and also the score. We also want to show the results of this analysis in an intuitive way by visualizing using histograms, word clouds, graphs, regression techniques and algorithms.

## 5. METHODOLOGY:

- Firstly we pick up confirmation from the twitter and youtube where tweets and comments can be extricated.
- There is a probability of extracting upto 1500 tweets from twitter. While youtube comments can be extracted in light of the comments of that show.
- Day insightful comments are extracted from youtube. While week astute and overall tweets are extricated from twitter.
- Here we take two reality shows and analyze which indicate is prevalent among them.
- Tweets and Comments of the two shows are extricated and cleaning of the tweets and commentsie. Emojis, URL and so forth ought to be evacuated.
- An addon included in the Youtube is we can ascertain the edges, vertices and degree of the comments.
- Lexical analysis ought to be performed ie ascertaining the over all score and rates of the positive and negative score .
- Information anlysis ie. producing diagrams, pie graphs, wordcloud ought to be performed.
- Linear and Logistic Regression can be performed on the score.
- Then a plot graph is used to compare the popularity of two shows

## 6.RESULT AND OBSERVATIONS:

### 6.1 Day wise Analysis Results:

a) Extraction of comments :

The table beneath demonstrates the comments got for Show 1 that is comedy nights with kapil and Show 2 as jabardasth. The comments were extracted from both the shows and are spoken to in table 1.

Show 1	Show 2
1. Show Server nice	1 Naresh Kumar Sirimella America ammaini pedadmu...puika
2. show parson by	2 simhachalam palli mana sampradayam lo battalu eskole...
3. pretty jye	3 satish kottapalli chudanu chudanu antune pratisari chu...
4. jeeba kadh for kadhney	4 satish kottapalli chudanu chudanu antune pratisari chu...
5. Appt error was without path... now what's the big deal	5 light RAY Islamic center nunchi vachava
6. idontia	
7. kadh jye	
8. Show langh	
9. Appt thama he kadh kadh bade he he	
10. the kadh thama die	
11. Nappi ?	

Table-1:Comments

b) Calculating Scores :

The table beneath demonstrates the scores got for Show 1 that is comedy nights with kapil and Show 2 as jabardasth. The scores were extracted from both the shows and are spoken to in table 2.

Show 1	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	neutral
1	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	1
5	0	1	0	0	1	0	1	1	0	1	0
6	0	0	0	0	0	0	0	0	0	0	1

Show 2	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	neutral
1	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0	1
11	0	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0	0	1
13	1	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0
16	-2	0	0	0	0	0	0	0	0	0	0

Table-2:Scores

### 6.2 Week wise Analysis Results:

a) Extraction of comments :

The table beneath demonstrates the comments got for Show 1 that is comedy nights with kapil and Show 2 as jabardasth. The comments were extracted from both the shows and are spoken to in table 3.

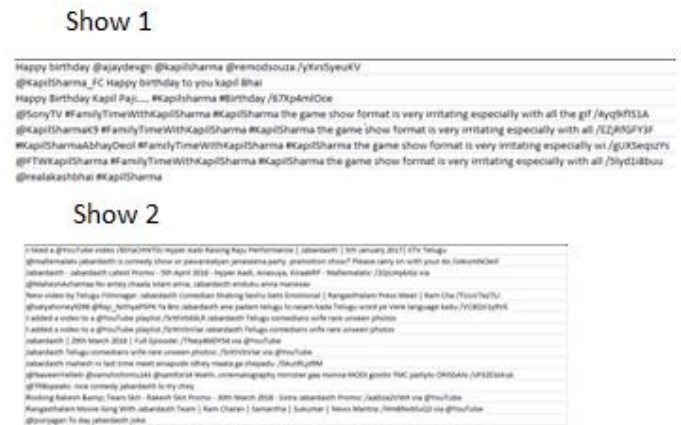


Table-3:Comments

b) Calculating Scores :

The table beneath demonstrates the scores got for Show 1 that is comedy nights with kapil and Show 2 as jabardasth. The scores were extracted from both the shows and are spoken to in table 4.

Show 1	Score	Positive	Negative	PosPercent	NegPercent
1	1	1	0	1	0
2	1	1	0	1	0
3	1	1	0	1	0
4	-1	0	1	0	1
5	-1	0	1	0	1
6	-1	0	1	0	1
7	-1	0	1	0	1
8	1	1	0	1	0

Show 2	Score	Positive	Negative	PosPercent	NegPercent
1	1	1	0	1	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	0	0	0	0	0
12	0	0	0	0	0
13	1	1	0	1	0
14	0	0	0	0	0
15	0	0	0	0	0
16	-2	0	2	0	1

Table-4:Score

### 6.3 Overall Analysis Results:

a) Extraction of comments :

The table beneath demonstrates the comments got for Show 1 that is comedy nights with kapil and Show 2 as jabardasth. The comments were extracted from both the shows and are spoken to in table 5.





Fig -4: Wordcloud

7.3 Overall Analysis Results:

a) Histograms:

The fig-5 beneath demonstrates the histograms got for Show 1 that is comedy nights with kapil and Show 2 as jabardasth. These histograms are created for positive scores, negative scores and the overall scores. From the fig-5 we can also observe that in show 1 there are more values ranging between the values -1 and 0 in score histogram whereas in show 2 there are more values ranging between 0 to -0.5 in score histogram.

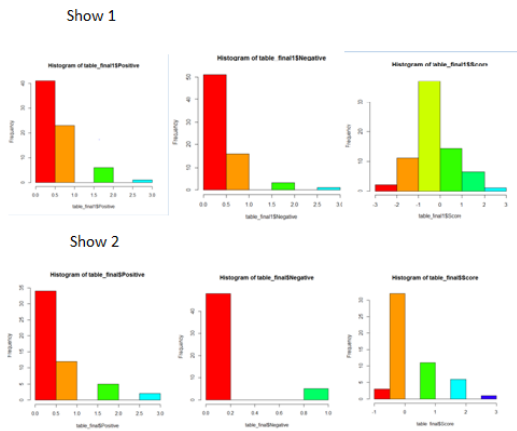


Fig -5: Histogram

b) Wordcloud:

Wordcloud in the fig-6 for both shows analyzes the comments and also quickly visualizes the keywords as a wordcloud.

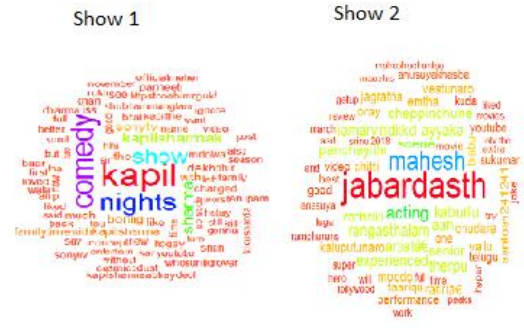


Fig -6: Wordcloud

8. REGRESSION TECHNIQUES :

8.1 Week wise Analysis Results:

a) Linear regression:

The point of linear regression is to display a constant variable Y as a scientific capacity of at least one X variable(s), with the goal that we can utilize this regression model to anticipate the Y when just the X is known. This scientific condition can be summed up as takes after:

$$Y = \beta_1 + \beta_2X + \epsilon$$

Where,  $\beta_1$  is the intercept and  $\beta_2$  is the slope. Collectively, they are called *regression coefficients*.  $\epsilon$  is the error term, the part of Y the regression model is unable to explain.

Output: The beneath fig-7 demonstrates that the score of show 1 that is 0.5 is more prominent than show 2 0.3 so in this way a conclusion can be made that the show 1 is famous.

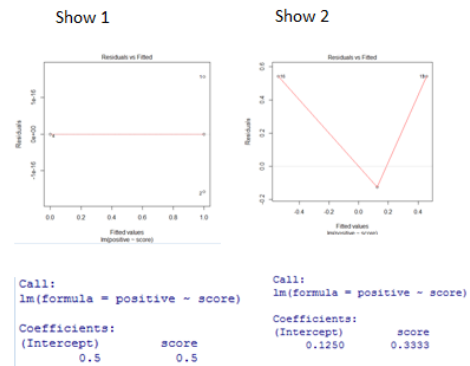


Fig -7: Linear Regression Graph.

b) Logistic regression:

In the linear regression serves to anticipate consistent Y factors, calculated regression is utilized for binary classification. In the event that we utilize linear regression to show a dichotomous variable (as Y), the subsequent model won't not limit the anticipated Ys inside 0 and 1. Moreover, other presumption of linear regression, for example, normality of errors may get disregarded.. So instead, we model the log odds of the event  $\ln(P1-P)\ln(P1-P)$ , where, P is the probability of event.

$$Zi = \ln(Pi1 - Pi) = \beta_0 + \beta_1x1 + \dots + \beta_nxn$$

The above equation can be demonstrated utilizing the glm() by setting the family contention to "binomial". So, the predicted values from the above model, i.e. the log odds of the event, can be converted to probability of event as follows:

$$Pi = 1 - (1 + e^{zi})^{-1}$$

Output: The beneath fig-8 demonstrates that the score of show 1 that is 0.5 is more prominent than show 2 0.3 so in this way a conclusion can be made that the show 1 is famous.

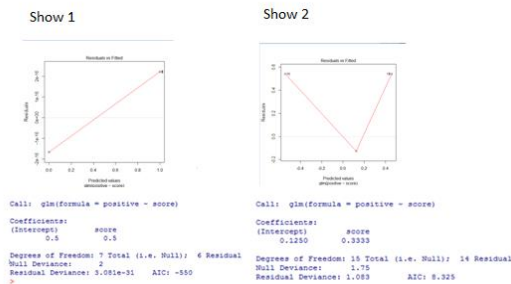


Fig -8: Logistic Regression Graph.

8.2 Overall Analysis Results:

a) Linear regression:

The point of linear regression is to display a constant variable Y as a scientific capacity of at least one X variable(s), with the goal that we can utilize this regression model to anticipate the Y when just the X is known. This scientific condition can be summed up as takes after:

$$Y = \beta_1 + \beta_2X + \epsilon$$

Where,  $\beta_1$  is the intercept and  $\beta_2$  is the slope. Collectively, they are called regression coefficients.  $\epsilon$  is the error term, the part of Y the regression model is unable to explain.

Output: The beneath fig-9 demonstrates that the score of show 1 that is 0.5 is more prominent than show 2 0.8 so in this way a conclusion can be made that the show 2 is famous.

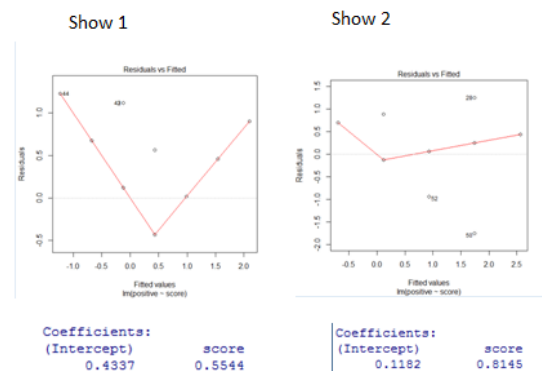


Fig -9: Linear Regression Graph.

b) Logistic regression:

In the linear regression serves to anticipate consistent Y factors, calculated regression is utilized for binary classification. In the event that we utilize linear regression to show a dichotomous variable (as Y), the subsequent model won't not limit the anticipated Ys inside 0 and 1. Moreover, other presumption of linear regression, for example, normality of errors may get disregarded.. So instead, we model the log odds of the event  $\ln(P1-P)\ln(P1-P)$ , where, P is the probability of event.

$$Zi = \ln(Pi1 - Pi) = \beta_0 + \beta_1x1 + \dots + \beta_nxn$$

The above equation can be demonstrated utilizing the glm() by setting the family contention to "binomial". So, the predicted values from the above model, i.e. the log odds of the event, can be converted to probability of event as follows:

$$Pi = 1 - (1 + e^{zi})^{-1}$$

Output: The beneath fig-10 demonstrates that the score of show 1 that is 0.5 is more prominent than show 2 0.8 so in this way a conclusion can be made that the show 2 is famous.

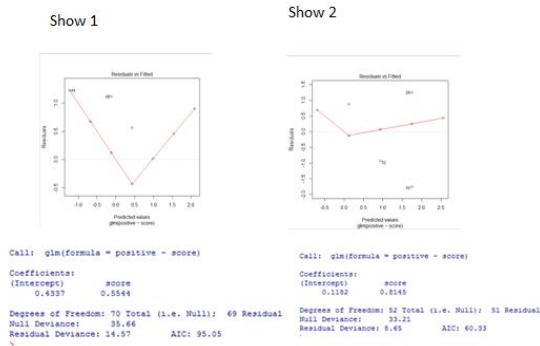


Fig -10: Logistic Regression Graph.

### 9. FINAL COMPARISON GRAPH :

The overall comparison of the two shows can be done by using the graph plot.

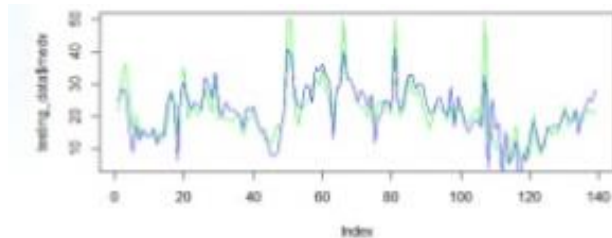


Fig -11: Comparison Graph.

The overall scores are taken and the comparison can be done as shown in the fig-11. The green line demonstrates the show 1 and blue line demonstrates show 2. By the fig-11 we can advise that as per the by overall scores the well known show is show 2.

### 10. CONCLUSION:

Here we display a framework TV Show Popularity Prediction utilizing Sentiment Analysis in Social Network for users, which foresee the fame of the show among a few shows, on-screen characters, reality shows and serials in view of in light of the content surveys which are getting from social networking websites like Twitter and Youtube. The benefits of utilizing this framework are that it helps in breaking down TV Show points of interest and rates expectation in light of Twitter tweets and YouTube comments.

### 11. ACKNOWLEDGEMENT

We are greatly honored that our respected guide and project coordinator Dr. Sunil Bhutada spent his valuable

time guiding us and persistently encouraging us throughout the duration of this project. We express our gratitude towards him for the helpful suggestions that improved the quality of this work. He has been there with us through every step of this project whenever we needed his help and advice. We extend our sincere thanks to our Head of Department Dr. V.V.S.S.S Balram for the cooperation throughout the course of this project. Lastly, we are thankful to all the teaching and non-teaching staff of the department of Information Technology and all those who directly or indirectly contributed to complete this work.

### 12. REFERENCES:

1. Tejaswi Kadam, Gaurav Saraf, Vikas Dewadkar and P. J. Chate, "TV Show Popularity Prediction using Sentiment Analysis in Social Network", International Research Journal of Engineering and Technology- Vol 4, Issue 11, Nov 2017.
2. Saraee, M, White, S and Eccleston, J "A data mining approach to analysis and prediction of movie ratings", International journal of Mechanical Engineering and Technology-Volume 9, Issue 1, Jan 2018.
3. 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- Volume 9, Issue 1, Jan 2018.
4. Priyank Jain, "A Novel Approach to Analysis of TV Shows using Social Media, Machine Learning and Big Data". International journal of Technological Exploration and Learning.