

Unfurling the Latest Patterns of Entertainment Consumption by Indian Audience: A Twitter Sentiment Analysis

Devesh Lowe, Bhavna Galhotra, Yukti Ahuja

Abstract: *It is explicit that Online Video programming has emerged on the Indian entertainment landscape and is pleasantly thriving on binge watching, the “anytime, anywhere” content. By far, Television has been the most watched and trusted media for Indian audience but the OTT media has given an alternative channel for entertainment. In this paper the authors investigate viewership metrics by adjudging opinion through Tweets and studying them qualitatively and quantitatively. Tweets have been captured for popular TV shows and trending web series in the year 2018. Results showcase growing fondness for web content and are indicative of web based video programming sustenance. The study is a first of its kind in India meant to explore the urban audience viewership pattern using Twitter Sentiment Analysis.*

Index Terms: *Television, online video program, web-based program, Twitter Sentiment Analysis*

I. INTRODUCTION

Entertainment in India is undergoing a transition, from the family ritual of watching TV together, to an individual, preference-based video content consumption. In January 2016, Netflix, an international video on demand (VOD) platform offering a library of TV series, documentaries and films, launched itself in India. Soon, the mainstream Indian TV channels launched their VOD platforms — Viacom18 launched Voot in March 2016; Are launched in April 2016. By December 2016, Netflix wasn't the only international platform in India. Amazon also launched Prime Videos in India with its bank of international TV shows and films.

Over the top (OTT) video consumption is a technology alternative that allows the replication of traditional home entertainment in a digital context, which can be consumed either on fixed broadband or mobile internet. The contemporary flavors of the content make it even more appealing. While these platforms intend to captivate the millennials, at the same time are cognizant of the fact that it would be herculean to make a dent in the existing monopoly of television content in the country. Indian households taste for television program consumption was built over content comprising submissive daughters-in-law, vamps, with everything extravagant and larger than life, also included cheesy romances and family dramas alongside the reality shows. The new kids on the bloc speculated that an audience, tired of the melodramatic sagas, would flock to VOD for its

sharp, realistic, international content. A report on the growth of digital media by Deloitte estimated that India had over 200 million online video viewers by the end of 2016. In fact, by the end of 2017, India had 462 million active internet users of which 430.3 are mobile internet users [1]. An offshoot of this trend is binge watching, arising consumption of online video content. According to the Indian Media and Entertainment Report 2017 by KPMG India and the Federation of Indian Chambers of Commerce and Industry (FICCI) [2], between 2016 and 2021 mobile video traffic in India is expected to grow at a compound annual growth rate of 68% and the number of video capable devices and connections is expected to grow 2.2 times, crossing 800 million in number. Video is expected to grow to 78% of the overall mobile data traffic by 2021. Also, on the budget side, if the amount of money spent on direct to home services along with movies at multiplexes is clubbed, the subscription of VOD platforms stands economical. The online streaming space is presently choc-a-bloc with service providers and competition is red hot. Players like Hotstar, Voot are wooing audiences with both free and pay by view TV programs and movies to gain a stronger foothold in the market.

Apart from making entertainment individually accessible, VOD platforms have enabled access to quality content. International shows can now be seen legally, and without censorship, Indian stories with more urban themes such as homosexuality, mafia wars, mature content as well as comedy have been able to reach their audience.

Television has been the go-to medium for entertainment since years until its virility was challenged by the newer formats with superior content, promising star cast and exceptional production.

Digital Consumer Survey stated that the percentage of consumers who preferred to watch programs on television sets tumbled by a significant 78% in the past one year [3]. This insight was powerful enough to motivate the researchers to explore the sentiments of viewers on television as well as online platforms. The same has been done through a powerful social media tool, Twitter. Now a days People can easily share the information, knowledge, thoughts and experiences with the world by using the social Media websites like blogs, forums, social networks, tweets and so on [4]. Due to its widespread popularity, there is a massive amount of user reviews or opinions produced and shared daily. Twitter being one of the most widely used social media micro blogging sites.

Revised Manuscript Received on December 22, 2018.

Devesh Lowe, Assistant Professor, JIMS Rohini Sector-5, Delhi
Bhavna Galhotra, Assistant Professor, JIMS Rohini Sector-5, Delhi
Yukti Ahuja, Assistant Professor, JIMS Rohini Sector-5, Delhi

In the current study we try to present a sneak peek into the viewership with the help of Twitter Sentiment Analysis capturing data for perception towards both-country’s top rated television programs and the alternate entertainment web-content platforms.

The work is divided into the following sections-Introduction, Methodology, Analysis and discussion followed by Conclusion.

II. LITERATURE REVIEW

a. Twitter as a tool: Umman, Diren and Ceimil(2017) [5]discussed the wealth of data generated by billions of tweets, shares, posts and discussionson social media every day. They dwell upon the significance of these platforms and on how they provide new opportunities for research about human behavior, information diffusion and influence propagation at a scale that is otherwise impossible. Morstatter and lieu (2017) [6] investigated data collection associated with social media and also proposed computational methods to assess biases due to the way a social media site makes its data available, to detect the bias from data samples without an access to the full data and also to mitigate bias by designing data collection strategies that maximize the coverage to minimize bias. The E marketers He, Zha and Li’s (2013) [7] described the in depth study which applies text mining to analyze unstructured text content on Facebook and Twitter sites of the largest pizza chains : Pizza Hut, Domino’s Pizza and Papa John,s Pizza. The result revealed the value of the social media competitive analysis and the power of text mining as an effective technique to extract business value from the vast amount of available social media data. (Ghiassi, Skinner and Zimbra, 2013) [8]Introduced an approach to supervise the feature reduction technique using n gram and statistical analysis to develop a twitter –specific lexicon for the sentiment analysis. They have augmented this Lexicon with the brand specific terms for brand related tweets. They show that the Twitter Specific Lexicon is significantly more effective in terms of classification and accuracy metrics.

b. Since the tool has become a valued resource for sentiment mining in fields of customer relation management, public opinion and text filtering, authors have used tweets from the twitter handle of the various web series and the TV programming series to gauge the public opinion.

c. Video programing and television: According to a study commissioned by the Interactive Advertising Bureau (trade group of digital publishers and media companies) conducted by research firm GfK there is change in the content people want to watch during the day time and the evening time as in the day time there is least content that is viewed by the customers but they want to get connected to the Local and national TV, also the TV sports channel.

Original digital video programming is attracting a growing audience—especially younger viewers 18-34 who are highly desirable to many advertisers [9]. Brands and media buyers need to be cognizant not only of the growth in popularity of original digital video, but also should recognize the variety of screens where this content is consumed.

YouTube was one of the digital media available to the consumers for watching the videos uploaded by the users which gained the YouTube views and also this gave advertisers, agencies and the TV programmers an opportunity

to leverage the connection between the TV viewers and digital views. This gave valuable means to communicate brand messages to the wider audience, and also grew the audience share.

III. METHODOLOGY

The study investigates the opinions of Indian viewers expressed on Twitter about popular television shows and online video programing. Twitter is one of the micro blogging site used by, many people to make the best choices based on others opinions, the biggest task to analyse and summarize the overall data. [10]. For gaining the public sentiment on Reality Shows and Family Drama aired in Hindi on Indian television channels, and as web series on on-subscription Internet based service providers, authors prepared a list of shows with highest popularity on Internet portals. For attaining the rankings of shows on both the platforms, these websites were most reliable owing to their followership and trustworthiness among the public in general.

They have been reflecting the public opinion for long enough time. The three portals were - MtWiki (BARC ratings), Google and Imdb. As described in Table 1 and Table 2, we compared three different websites for most popular television shows aired on Indian television for Hindi audience in 2018.

To collect data for online video programing the researchers shortlisted web series, since they offered original content. To arrive at the most popular web series researchers collected data from their web sites, but as only a limited number of shows were web-casted, we selected all those which showed more than 100 tweets using their hash tags (Table-3).

Table 1: Rankings of Most popular drama shows [11], [12]

S.No	MtWiki (BARC ratings)	Google	imdb
1	Naagin	YehRishta Kya kehlatatahai	Naagin
2	KundliBhagya	Kumkumbhagya	bepannaah
3	Kumkumbhagya	Tarak Mehta kaooltaChashma	Tarak Mehta kaooltaChashma
4	Kulfi Kumar Bajewala	Kumkumbhagya	silsilabadalterishtonka
5	YehRishta Kya kehlatatahai	Naagin	kumkumbhagya

Table 2: Rankings of Most popular Reality shows [11], [12]

S.No	MtWiki (BARC ratings)	Google	Imdb
1	Indian Idol	BiggBoss	BiggBoss
2	KBC	Indian Idol	Dance Diwane
3	Dance Plus 4	India's Got Talent	JhalakDihlaJa
4	BiggBoss 12	DuskaDum	Dance Plus 4
5	IndiasGot Talent	Dance India Dance	Devlok with DeyduttPatnaik



Table-3: List of popular web-series

Title	Platform	Genre
Breathe	WebSeries	Drama
Comicstaan	WebSeries	Reality Show
Mirzapur	WebSeries	Drama
Sacred Games	WebSeries	Drama
Inside Edge	WebSeries	Drama
HearMeLoveMe	WebSeries	Reality Show

One limitation of the study is that it could fetch the data of only those shows which either had their own Twitter handle or used hashtags. It was disappointing to see that many of the shows listed in table1 and table2 were not using a formal twitter handle and were not very popular on twitter with their own hashtags, hence the researchers were not able to gather data on #KBC and #Naagin3. Following steps were taken for data collection and analysis. Below the Fig 1 Shows the complete analysis process from the data collection to the outcome of the Sentiments presented in the analysis section.

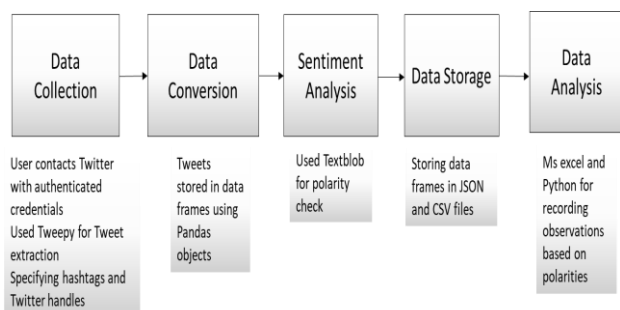


Figure-1 Depicting the functionality of the Model on Tweets (created by authors)

Data Collection: authors collected data from well-known social media platform Twitter. Twitter is extremely popular with common users for voicing their opinion about any current social, political or personal issues. Twitter data is available for developers from developer.twitter.com by creating a developer account. Authors created a developer account with twitter by using consumer key, consumer secret key, access token and access secret key. Authors used **Python 2.7** with ‘**tweepy**’ library for extracting data from twitter. Data was retrieved using python scripts in **anaconda** framework using **Twitter API**. Data was collected from tweets generated in 2018 using twitter handles and #tags related to television series and web series aired in 2018 on Indian TV channels and web portals. Scripts used were as follows:

```

auth = tweepy.OAuthHandler (CONSUMER_KEY,
CONSUMER_SECRET)
auth.set_access_token (ACCESS_TOKEN,
ACCESS_SECRET)
api= tweepy.API(auth)
first200 = api.user_timeline(screen_name=scr_nam,
count=200)
alltweets=get_tweets(first200,scr_nam,first200[-1].id)
In the above code we used scr_nam to hold screen name or
twitter handle for TV series. Since tweets retrieved at one
time was limited to a fix value, a separate function call
get_tweets() was created to use api.user_timeline() method
iteratively in order to read multiple tweets.
  
```

Data conversion: All data hence extracted was stored in ‘**pandas**’ object with separate columns describing different features of every tweet. Benefit of storing data in pandas object was that it could be easily converted to csv file later.

```

data = pd.DataFrame(data=[tweet.text for tweet in
alltweets], columns=['Tweets'])
  
```

Similarly, separate features were created in pandas object ‘pd’ to hold various attributes with proper titles.

Sentiment Analysis: authors used ‘**textblob**’ library for sentiment analysis on every tweet. Sentiment analysis on tweets is done using word token library in textblob which identifies the polarity of tweets on the basis of words used. Polarity of every tweet is identified as

Positive Tweets: tweets which convey positive sentiment by use of positive words like good, like, best, wonderful, beautiful, wow etc.

Negative Tweets: tweets which convey negative sentiment by sue of negative words like not, don’t, bad, worse, avoid etc.

Neutral tweets: tweets which do not convey any sentiment.

data['SA'] = np.array([analyze_sentiment(tweet) for tweet in data['Tweets']]) to execute sentiment analysis for every tweet. we created a separate function analyze sentiment() which uses a python script to analyze the polarity of tweet and store it in separate feature in data object.

Sentiment analysis was done by checking the sentiment polarity of every tweet using textblob.sentiment.polarity(). Textblob is a library in Python 2 for checking textual data and it provides a simple API for NLP. Textblob calculates polarity of every phrase by using lexical analysis. For example a simple statement textblob(“a wonderful day”).sentiment will provide us with polarity and subjectivity of the statement mentioned in parenthesis. Polarity reflects the positivity or negativity of lexical unit as recorded in library (ref). Subjectivity refers to the relevance of units in the phrase. We have focused on polarity of tweets for developing this paper. Polarity >0 is considered positive and polarity <0 is considered negative. Tweets with Polarity =0 are considered as neutral tweets.

Data storage: all data collected along with their sentiment analysis is stored in separate .csv file for data analysis.

```

data.to_csv("IGT.csv", encoding='utf-8', index=False)
  
```

Data Analysis: since we collected data for more than 20,000 tweets on TV shows and Web-series and saved it in .csv files, we started analyzing data using Python and MS-Excel.

3.1 DATA ANALYSIS, DISCUSSION AND CONCLUSION

Data was collected till first week of December 2018 and authors present an analysis of the same.

A. Popularity of TV and Web-series: authors observed that Indian viewer is frequently tweeting about television shows and web-series. Though authors acknowledge the reach and penetration of TV shows in Indian audience, but also observed that popularity of TV shows and Web series is at par. With the availability of high speed internet on very low cost and increased use of smart phones,



Indian urban audience is regularly watching web-series and have shown good interest in them. As shown in Table 4, number of tweets for most popular TV shows and web-series is almost equal.

As recorded above, most popular shows aired on television and Internet channels, both recorded similar number of tweets. Web-series titled 'Breathe' aired on Amazon Prime Video recorded 3200 tweets appx. With its #tag or handle mentioned in them. Similarly most popular shows on television 'KumKumBhagya' also recorded 3299 tweets. Also to note that most popular reality shows on television like 'BiggBoss12' and 'Indian Idol' also got their name mentioned in tweets which is comparable to reality show 'comicstaan' aired on Prime videos. This clearly indicates that popularity of TV shows and Web series is at par in both Family Drama category and Reality Show category.

B. Positivity in tweets for Family Drama and comedy shows: Authors also recorded that percentage of positive tweets for family drama category or comedy / entertainment shows is much higher than, positivity in tweets for shows which serve violence/fights/aggression/mature/adult content. We observed the same trend in both family drama category and reality show category. This trend is also followed in shows webcasted through internet (figure 2 and 3).

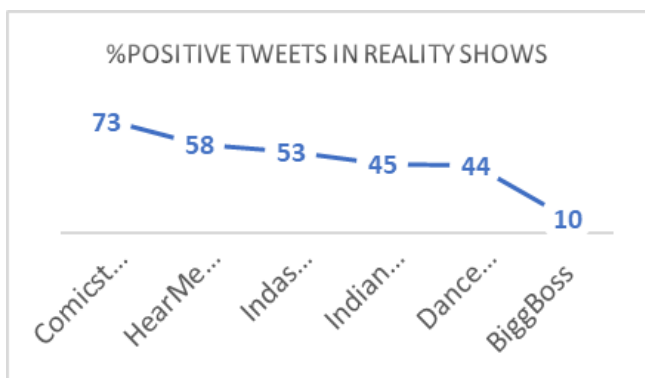


Figure 2: percentage of positive tweets in Reality shows

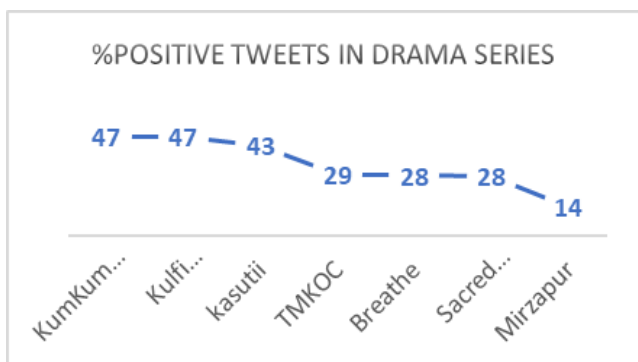


Figure 3: Percentage of positive tweets in Drama series

As depicted in figure 2 and figure 3, reality shows like 'India's Got talent', 'Indian Idol' aired on television, and 'comicstaan' aired on prime videos, received huge response from twiterati. But shows like 'BIGGBOSS', although received huge response but did not receive positive tweets. Similar observation was recorded for drama shows. Most popular shows aired on television like 'KumkumBhagya' and 'Kulfi Kumar Bajewala' not only received huge response in

form of tweets, but also a lot of positive tweets. On the other hand shows like 'Mirzapur', 'Breathe' and 'Sacred Games' did not receive positive tweets.

C. Web-series engage viewers but only during its web-cast. A popular belief is that web-series can be downloaded and can be viewed later, which increases its viewership. Authors' data and analysis show that most popular web-series are only popular when they are web-casted. Within that time-frame, viewers view this content and tweet about it, but are not responding to its twitter handle afterwards. As shown in figure 4 and figure 5 below, most popular drama series and most popular reality series were not popular throughout the year.

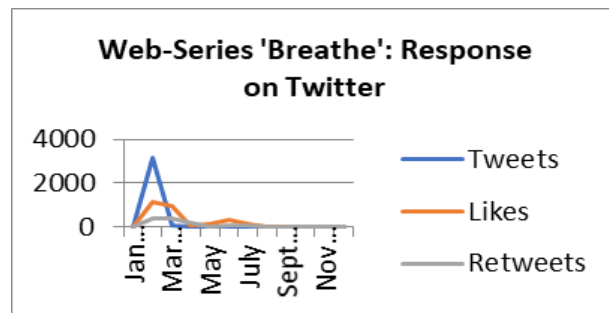


Figure 4: Timeline Response for Breathe

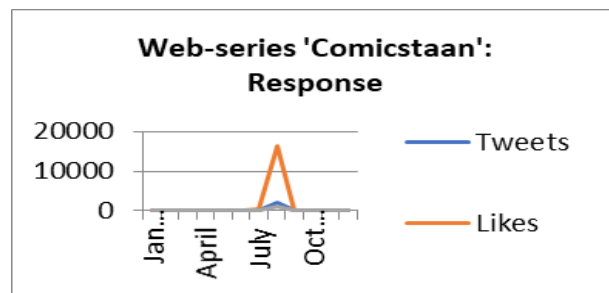


Figure 5: Timeline response for Comicstaan

D. Web-series as entertainment platform: recognized and at par with TV audience. Our comparison drawn in Table 5 below explicitly shows the response of Indian audience towards TV shows and web-series. People are not only accepting web-series as an alternate form of entertainment but also are regularly voicing their opinion in form of tweets, retweets and likes. The table shows that the number of "likes" used for TV series is way above the likes used for Web series. This does reflect highly upon the endorsement for TV series content but at the same time hinges at reluctance to put their opinion by Tweeting. The percentage of positive tweets on Web series are 45% whereas on television content it is 33%.



Table 5: Comparison of Tweets on TV-content and Web-series

<u>Platform</u>	<u>Total tweets</u>	<u>Positive tweets</u>	<u>Likes</u>	<u>%age of Positive Tweets</u>
Web Series Total	5767	2624	69355	45.5002601
TV Series Total	9167	3045	307720	33.2169739

III. CONCLUSION

Just a year into this digital revolution, it's too early to predict how VOD platforms will change the quality of alternate content but it surely has made inroads into the Indian entertainment consumption. Ditching television will not be easy and embracing the new service will have to be taken with a pinch of salt .Netflix and Amazon could have an advantage in India because of their deep pockets and global expertise. Superior quality, content and affordability will remain to be a promising strategy. Since the patterns may take a little while to change, a valid question is whether companies will stay invested patiently. Researchers recommend a qualitative analysis of data in order to study the opinions in depth and also to take strategic decisions accordingly.

REFERENCES

1. "Statista," January 2018. [Online]. Available: <https://www.statista.com/statistics/309866/india-digital-population/>. [Accessed 16 November 2018].
2. "KPMG India," 21 march 2017. [Online]. Available: <https://home.kpmg.com/in/en/home/events/2017/03/kpmg-india-ficci-media-entertainment-report-2017.html>. [Accessed 12 october 2018].
3. E. B. Matteo Maga, "Channelpostmea," 2017. [Online]. Available: http://www.channelpostmea.com/wp-content/uploads/2017/08/Accenture_DCS_UAE-Infographic-highlights_final.pdf. [Accessed 8 november 2018].
4. A. Agarwal and J. S. Sabharwal, "End to end Sentiment analysis of Twitter Data," Proceedings of Workshop on information and Entity analytics, pp. 39-44, 2012.
5. D. B. Y. Umman Tugba Gürsoy, "Social Media Mining and Sentiment Analysis for Brand Management," Global Journal of Emerging Trends in e-Business, Marketing and Consumer Psychology (GJETeMCP), vol. 3 , no. 1 1, pp. 497-511, 2017.
6. J. P. L. M. C. Fred Morstatter, "Is the Sample Good Enough? Comparing Data," Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media , 2013.
7. Social media competitive analysis and text mining: A case study in the pizza industry, 2013.
8. D. Z. S. M. Ghiassi a, "Twitter brand sentiment analysis: A hybrid system using n-gram," Expert Systems with Applications, 2013.
9. "Sherrill Mane Recognized in Folio: Top Women in Media," IAB, 2016.
10. B. Liu, web data mining : Exploring Hyperlinks Contents , and usage data, Springer, 2006.
11. "Mtwiki," 13 December 2018. [Online]. Available: <https://mtwiki.blogspot.com/2018/05/top-10-indian-tv-serials-by-trp-rating.html>. [Accessed December 2018].
12. "Imdb," october 2018. [Online]. Available: <https://www.imdb.com/list/ls040034559/>. [Accessed 2 December 2018].
13. F. F. G. G. Alessia D'Andrea, "Approaches,Tools and Applications for Sentiment Analysis Implementation," International Journna of Computer applications, vol. 125(3), no. 10.5120/ijca2015905866, pp. 26-33, 2015.

14. M. M. ., K. ., K. Miljana Mitic, "Upcovering customer service experiences with Twitter: the case sudy of airline Industry," Emerald, vol. 52, no. 4, pp. 705-723, 2014.
15. S. V. Bo Pang and Lillian Lee, "Thumb up ? Sentient Classification using Machine Learning techniques," Empirical methods oin Natural Language Processing ,Association for Computational Linguistics, pp. 79-86., 2002.
16. S. W. G. Kiruthika M, "Sentiment analysis of Twitter Data," IJiet, vol. 6, no. 4, pp. 264-273, 2016.
17. Rachel, Twitter Sentiment analysis for Marketing Research, 2014.
18. "Chicago Triune," Tribune news services, 11 May 2016. [Online]. Available: <https://www.chicagotribune.com/entertainment/tv/ct-digital-video-vs-p-rimetime-tv-20160511-story.html>. [Accessed 12 november 2018].
19. A. CORSELLI, "Dmnews," 12 MAY 2015. [Online]. Available: <https://www.dmnews.com/channel-marketing/news/13056021/original-digital-video-programming-on-the-rise>. [Accessed 14 November 2018].
20. The Relationship of Digital Video Programming Viewership And Tv Programming, 2016.
21. "Knowledge.Whatron," Whatron, 4 May 2017. [Online]. Available: <http://knowledge.wharton.upenn.edu/article/can-netflix-amazon-disrupt-indias-streaming-video-market/>. [Accessed 2 November 2018].
22. P. Neogy, "Yahoo Finance," 1 may 2017. [Online]. Available: <https://in.finance.yahoo.com/news/understanding-crowded-video-demand-vod-market-india-174629027.html?guccounter=1>. [Accessed 14 October 2018].

