iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XII **Month of publication:** December 2017

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Online Word of Mouth Communication in Bollywood Tweet Dataset

Sandeep Ranjan[1], Dr. Sumesh Sood[2]
[1]PhD Scholar, I.K. Gujral Punjab Technical University, Kapurthala, India
[2]Assistant Professor, I.K. Gujral Punjab Technical University, Kapurthala,India

Abstract: Social networks are becoming a popular place to share opinions about latest happenings such as a new movie release. Movie's success depends mainly on the first week box office collection and requires fast advertisement with broad coverage. Twitter serves this purpose by spreading online word of mouth, which acts as a stimulus to brand promotion. Users read other users tweets and express their own opinion about the topic. In this research, tweets for Twitter handles of Bollywood Hindi movies released between 1st June 2017 and 28th July 2017 were fetched to identify top word pairs mentioned by users. This helps to understand the word of mouth based information spread in online social networks and identify most active nodes and edges in the graphs.

***Keywords*: brand promotion, graphs, online word of mouth, social networks**

I. INTRODUCTION

Word of mouth (WOM) has been defined as an informal communication among customers about services and products [1]. It ranges from casual talks between two persons to promoting a particular brand focusing an entire population domain. Advancement in networking and Internet technologies has helped WOM creep into electronic media communication [2]. Social network websites, discussion forums, blogs and news groups are live examples of online word of mouth. Web users share their views and opinions related to events, movies, music albums, news, and stock prices with other users who are completely unknown to them[3]. Users express positive, negative or neutral views or support somebody else's views by liking or sharing it on their online profile.

Companies use online word of mouth (OWOM)as a powerful consumer to consumer influence tool backed by marketing professionals [4]. Today, most of the corporations, service providers, organizations and even government bodies have their presence on social networking websites to increase their influence masses or to get a feedback of their activities in order to make future policies and course of action. Companies use social networking media to collect and solve customer complaints[5]. Twitter has been aggressively used in campaigning by the contestants in various Presidential and Parliament elections [6], [7]. Evidently OWOM has helped politicians reach the masses and frame strategies to win elections by gathering feedback, do timely analysis of their public brand value and take corrective actions.

Word of mouth communication is a popular form of online user interactions primarily within the communities[8]. Communities are formed of connected vertices; and vertices get connected when they exchange information with each other[9]. Thus, communities and OWOM go hand in hand. A denser community having a large number of constituent vertices is an essential criterion for the spread of OWOM.

II. LITERATURE REVIEW

Social network websites have got a major role in brand building and consumer relationship .Researchers analyzed #McDStoriest wets which erupted with negative comments from users, forcing McDonalds' to withdraw the promotion event within just 2 hours of launch on January 18, 2012[11]. This signifies what companies have to go through when a large number of customers get connected and share their experiences leading to a chain reaction of either positive or negative sentiments. The Bollywood actor network was studied a weighted real world network to reveal the role of individual nodes in determining the overall outcome of sub graph communities of the network [12]. The community detection and analysis feature of social networks treated as graph helps to understand the relation between mutual funds and constituent stocks[13]. It was investigated that how the collected opinion of the general public correlates with the Dow Jones Indices [14]. The research shows 87% accuracy in daily closing values, up and down change prediction ofDow Jones. Stories and posts which are of great interest and relevance gather attention of the masses and try to influence readers with a repository of already shared opinions of responders. This starts information chain building as new users keep on adding to the network by sharing, re-tweeting and lining the posted content. Twitter data have been used to predict online music sales [15], [16]. Predicting well in advance whether a new artist would be a good bet can be of great use to producers and managers. Posting about new music release and analyzing user feedback gives a good hintto the studio about the possible sales figure that will be generated in coming weeks.

Companies like MasterCard, Google Federal Express and many others are providing real-time user activity data about different products and services which can be used to predict future trends and sales [17]. Millions of users use such economic platforms while making online purchases, any patterns discovered in these search graphs can be used to influence and other customers and converting their search into a successful transaction. Twitter has Promoted Tweets service for buying promotional tweets from brands to be shown to users just like regular tweets in addition to the brand pages followed by the users [18]. Presently, most of the companies are active on Twitter and they tweet their brand posts to the users who are following them. This is a quick and economical method to exchange information and also forms the basis of the data analysis. This model of EWOM is successful as more than 86% companies are following it. Social media has been used as a competitive analysis tool by businesses[19]. EWOM data of a large number of users is freely available on Twitter. Companies require monitoring of content generated by the customers about their product to stay ahead in business or at least stay in business. The collective wisdom of the masses can be used forecast business outcomes and improvement in their services[20]–[23]. The degree of a network node plays an important role in both gathering and disseminating of information through EWOM. The degree can be in degree or out degree, but it simply means how many other nodes are connected to a particular node. Higher the connectivity more will be information exchange due to the number of connected neighbors.

Social networks enable making friends beyond regional and country boundaries and also allow users to follow others, thus help to track likes and activities of the one being followed. In social networks, friends are an important influencing factor in consumer purchasing trends [24]. Facebook and Twitter update users of their friend's brand likes and the user is prompted to click on the brand page or post thereby increasing the probability of turning them into a consumer of that product or service. Certain keywords also give a similar stimulus to users by drawing their attention. Twitter hashtags mentioning earthquakes, cyclones, discount offers, sports events and celebrity news get instant user attention with the help of OWOM the moment initial news about them is announced. OWOM has the capability to influence sales and perceptions and is drawing attention of scholars [25]. OWOM was collected from two websites www.edmunds.com and www.consumerreports.org was analyzed and compared with sales figures of the automobile sector.

III.DATASET TIMELINE

The Bollywood tweet dataset was created by fetching Tweets for Twitter handles of Bollywood Hindi movies released between 1st June 2017 and 28th July 2017 using Node XL pluginfor MS Excel[26], [27]. Bollywood Hindi movies try to earn the most revenue in the first week as there are a number of movies piled up waiting to be released in the coming weeks. First week tweets (six days from the release) for a total of 22 movies were collected. Only distinct tweets were selected for the process by applying filter to Tweet column. The number of connected components was calculated using the 'Group by Connected Component' feature of NodeXL.

TABLE I: Dataet Description

S. No	Movie Name	Release Date	Twitter handle	Total Tweets	Total Distinct Tweets	No of connected components
1	Sweetie Weds NRI	02/06/17	sweetiewedsnri	2560	138	23
2	Mirror Game	02/06/17	mirrorgamefilm	39420	678	23
3	Hanuman da Damdadaar	02/06/17	hanumandadamdaar	23705	1269	104
4	Flat 211	02/06/17	flat211	3682	119	14
5	Dobaara	02/06/17	dobaara	11392	2453	462
6	Dear Maya	02/06/17	dearmaya	20608	1312	198
7	BehenHogi Teri	09/06/17	behenhogiteri	17884	1989	252
8	Raabta	09/06/17	raabta	14856	4391	1250
9	Love You Family	09/06/17	loveyoufamily	227	24	10
10	Bank Chor	16/06/17	bankchor	8718	1747	305
11	Tubelight	23/06/17	tubelight	14780	4657	924
12	EkHaseenaThiEkDeewana Tha	30/06/17	ehtedt	11015	932	23
13	Mom	07/07/17	mommovie	1200	59	30

14	Guest In London	07/07/17	guestinlondon	4657	383	97
15	JaggaJasoos	14/07/17	jaggajasoos	13781	5105	672
16	Shab	14/07/17	shabthefilm	7359	970	25
17	Lipstick Under My Burkha	21/07/17	lipstickundermyburkha	30079	3360	874
18	Munna Michael	21/07/17	munnamichael	54494	3720	427
19	RaagDesh	28/07/17	raagdesh	16653	1362	216
20	InduSarkar	28/07/17	indusarkar	43663	3123	653
21	Mubarakan	28/07/17	mubarakan	25761	2463	780
22	Baarat Company	28/07/17	baaratcompany	1337	54	15

Table 1 describes the dataset in terms of movie Twitter hashtags, release date, total number of tweets fetched during the first week of release, number of distinct tweets after obtained applying the filter and the number of connected components in the graph.

IV. EXPERIMENT FRAMEWORK

The dataset created by fetching tweets for movies over a period of two months was analyzed using NodeXL. NodeXL has some very useful features like ‘Words and word pairs’ and ‘Twitter search network top items’. They are available under the Graph Metrics option under the NodeXL tab. When the check boxes of these tools are clicked under the Graph Metrics, ‘Tweets’ column of the worksheet is to be selected. Now the top word pairs, top mentioned to and top replied to parameters will be computed from the ‘Tweets’ column. Table 2 contains the values for these parameters for the entire graph dataset of ‘Sweetie Weds NRI’ movie.

TABLE III
‘SWEETIE WEDS NRI’ MOVIE WORD COUNT ANALYSIS

Top Word Pairs in Tweet in Entire Graph	Entire Graph Count
himanshkohli,zoyaafroz	26
sweetiewedsnri,himanshkohli	20
promote,upcoming	18
himanshkohli,amp	16
amp,zoyaafroz	14
zoyaafroz,promote	14
upcoming,film	14
film,sweetiewedsnri	14
sweetiewedsnri,delhi	14
movie,sweetiewedsnri	11
Top Replied-To in Entire Graph	Entire Graph Count
himanshkohli	29
sheraya_ghosh	1
cyrusdastur	1
Top Mentioned in Entire Graph	Entire Graph Count
himanshkohli	61
zoyaafroz	45
kingkohlifc	13
kingkohli_team	8
palash_muchhal	5
farheen027	3
dilseradio	2
krazay_niki	2
arkopravo19	2

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

nishalovehk	2

The 'top word pairs' parameter tells which word pairs occurred together in the entire tweet dataset. These are the words which formed most of the tweet content from the movie's Twitter handle. In the case of 'Sweetie Weds NRI' movie, the most common word pair is himanshkohli-zoyaafroz which occurred 26 times. himanshkohli and zoyaafroz are the Twitter usernames of the lead actors of this movie:-HimanshKohli and ZoyaAfroz. Their usernames are the top replied to and top mentioned in the entire dataset. In other words, it can be phrased that these usernames started the initial conversation on Twitter for their movie promotion and remained most replied to and most mentioned in the dataset consisting of the general public tweets during the given period. They started the OWOM and helped it spread across the masses.

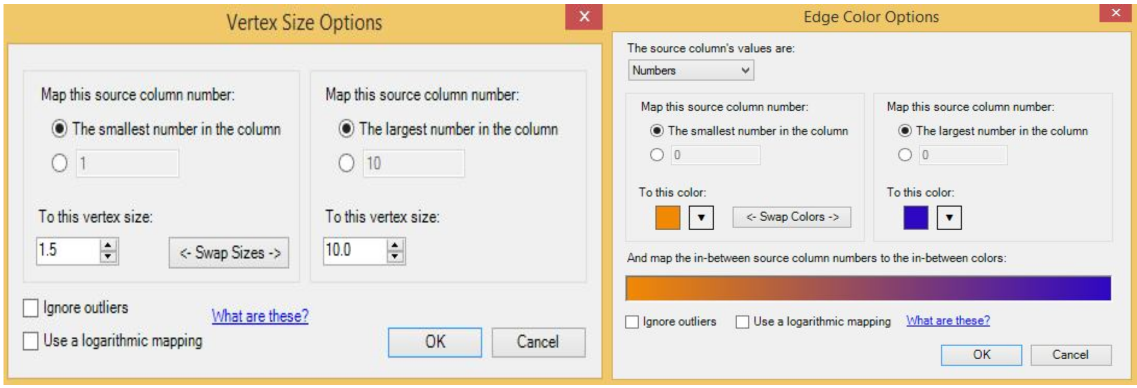
TABLE IIIII
'SWEETIE WEDS NRI' MOVIE VISUAL PROPERTIES AND GRAPH METRICS

	Visual Properties		Graph Metrics		
Vertex	Color	Size	Degree	In-Degree	Out-Degree
himanshkohli	111, 51, 131	7.2	4	2	2
zoyaafroz	111, 51, 131	7.2	4	2	1
sweetiewedsnri	46, 7, 195	10.0	6	3	2
promote	176, 93, 68	4.3	2	1	1
upcoming	176, 93, 68	4.3	2	1	1
amp	176, 93, 68	4.3	2	1	1
film	176, 93, 68	4.3	2	1	1
delhi	176, 93, 68	4.3	2	1	1
movie	241, 137, 4	1.5	0	0	1
rt	241, 137, 4	1.5	0	0	2
kingkohlifc	176, 93, 68	4.3	2	1	0
2nd	176, 93, 68	4.3	2	1	1
june	176, 93, 68	4.3	2	1	0
releasing	176, 93, 68	4.3	2	1	1
bollywood	241, 137, 4	1.5	0	0	1
stars	176, 93, 68	4.3	2	1	1
sweetie	241, 137, 4	1.5	0	0	1
weds	176, 93, 68	4.3	2	1	1
song	241, 137, 4	1.5	0	0	1
kingkohli_team	176, 93, 68	4.3	2	1	0
nri	176, 93, 68	4.3	2	1	0
lots	241, 137, 4	1.5	0	0	1
love	176, 93, 68	4.3	2	1	0

A graph was created using NodeXL to plot the top word pairs having vertices labeled by the first word of the pair. The size and color of each vertex are dependent on the degree of the vertex (number of nodes directly connected to that node).The opacity of edges is also determined by the number of pairings of the words which are connected by that edge. Edge opacity increases with the number of pairings of the words and so do the edge colors vary.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Figure 1 shows the variation of color for both the edges and vertices depending upon the degree of vertices. The color range can be selected from the available colors. Similarly the vertex size and edge opacity can be obtained by specifying the range in respective options under the NodeXL 'Autofill columns'.



(a) NodeXL Vertex Size Options(b) NodeXL Color Options

Fig. 1 Node XL Autofill column options

From the list of top word pairs in Table 2, the pairs occurring at least five times along with their degree were selected and pasted into a new NodeXL worksheet. Labels, edge color options, vertex options and edge opacity parameters were applied to obtain the following network graph.

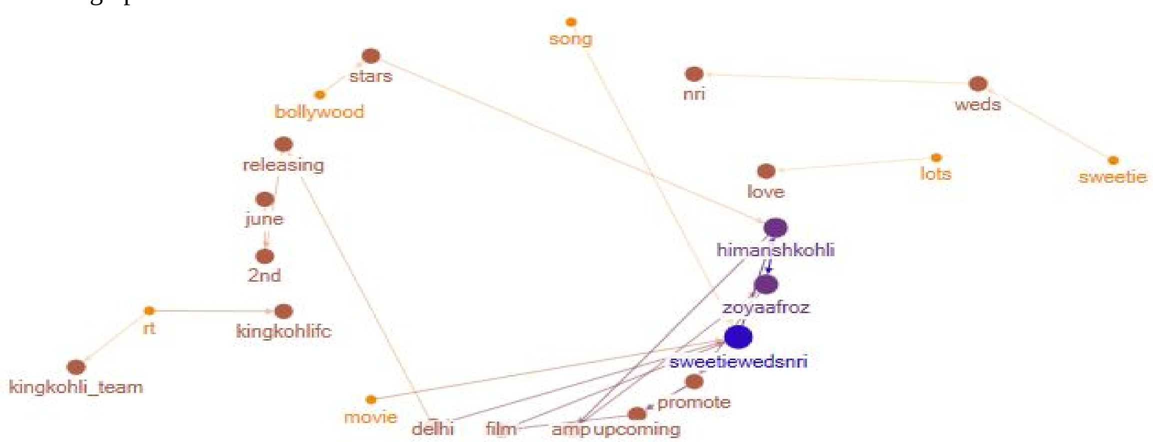


Fig. 2 Graphical analysis of 'Sweetie Weds NRI' movie top word pairs

Figure 2 shows the graph of 'Sweetie Weds NRI' movie top word pairs. Evidently the node for the word 'sweetiewedsnri' having highest degree has the largest node size and brightest color for the selected range. The groups of vertices kingkohli_team-rt-kingkohlifc, nri-weds-sweetie and love-lots are isolated from the main bigger connected component. This depicts the spread of OWOM in this network. Similarly, different networks an be analyzed for the spread of OWOM. This helps in brand monitoring and getting timely feedback to plan strategies for improving the results.

TABLE IVV
TOP WORD PAIRS AND COUNT

Movie	Top word pairs	Count	Movie	Top word pairs	Count	Movie	Top word pairs	Count
sweetiewedsnri	himanshkohli-zoyaafroz	26	mirrorgame	mirror- game	208	dobaara	releases,today	928
	sweetiewedsnri-himanshkohli	20		game- film	197		dobaara,releases	916
	promote-upcoming	18		film-mirrorgamefilm	148		humasqureshi,saqibsaleem	845
	himanshkohli-amp	16		mirrorgamefilm-iampoojabatra	138		theofficialb4u,humasqureshi	748

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

	amp- zoyaafroz	14		koinikalpadahai-fipb	102		today,theofficialb4u	584
hanumandadamdaar	hanuman-damdaar	457	flat211	2nd- june	33	dearmaya	mkoirala,dearmaya	78
	being-salmankhan	171		flat211- song	26		film,dearmaya	65
	Taher- 07	53		june- release	26		review,dearmaya	58
	Ruchinarain- 18	52		release-md_irfan17	26		manisha,koirala	55
	Bookmyshow-movie	33		md_irfan17-aslidivyakumar	26		watch,dearmaya	49
behenhogiteri	muktaa2cinemas- contestalert	200	raabta	box- office	487	loveyoufamily	film,'	2
	contestalert-behenhogiteri	147		sushant- singh	409		',loveyoufamily'	2
	rajkummarrao-shrutihaasan	130		singh- rajput	366		loveyoufamily',shaktikapoor	2
	rt- oddballindia	102		itsssr- kritisanon	342		shaktikapoor,akshapardasany	2
	behenhogiteri-cinemas	102		kriti- sanon	264		akshapardasany, manoj	2
bankchor	review-bankchor	167	tubelight	salman- khan	250	ehtedt	ehtedt,shivsdarshan	398
	riteishd-vivek_oberoi	157		box- office	154		shivsdarshan,ehtedt	236
	https- t	112		https- t	150		ekhaseenathiekdewanatha,ehtedt	202
	t- co	112		t- co	149		ek,haseena	179
	5- review	107		watching-tubelight	125		haseena,thi	179
mommovie	showing- power	6	guestinlondon	movie-guestinlondon	31	jaggajasoos	ranbir- kapoor	265
	power- mom	6		watched-guestinlondon	26		jagga- jasoos	179
	sridevibkapoor-zee	6		theaaryankartik-kriti_official	23		ranbirkapoor-katrinakaif	177
	zee- bangla	6		guestinlondon-movie	13		box- office	153
	bangla- show	6		guestinlondon-theaaryankartik	12		watch-jaggajasoos	148
shabthefilm	_pvrcinemas-shabthefilm	108	lipstickundermyburkha	box- office	193	munnamichael	box- office	267
	shabthefilm- pvr	80		1- 22	141		itigershroff-agerwalnidhhi	177
	iamonir-sanjaysuri	69		2- 17	118		munnamichael-itigershroff	163
	shabthefilm-iamonir	67		fri- 1	116		munna- michael	139
	iamonir-shabthefilm	58		watch-lipstickundermyburkha	114		office- collection	138
raagdesh	screening-raagdesh	129	indusarkar	watch- indusarkar	137	mubarakan	box- office	109

	special-screening	126		indusarkar- movie	116		anilkapoor-arjunk26	105
	varun- dhawan	70		review- indusarkar	105		arjunk26-anilkapoor	100
	raagdesh- movie	68		releasing-indusarkar	100		panamaverdict-mubarakan	92
	raagdesh-screening	61		review- movie	99		arjun- kapoor	79
baaratcompany	thefreejinn-contest_hub	9						
	contest_hub-contest_in	9						
	baaraatcompany-iamsandeepadhar	9						
	contest_in-contestmantra	9						
	iamsandeepadhar- afzalistan	8						

Table 4 shows the top word pairs for all movie tweet datasets. These are the words which act as most active nodes in the network graphs. The majority of the conversation in the dataset involved these words as they refer to the main references of the movie, its star cast, movie plot or the theme.

V. CONCLUSION

The popularity of social networks like Twitter gives an opportunity to brand owners and promoters to use this media as a cost effective and real time means of promotion and monitoring. Popular events like Bollywood movies get their promotion from OWOM through social network users. Real time analysis of tweet datasets can help identify most common word pairs from the tweet content, giving an idea what is the general public talking about. These words are, either positive or negative publicity about the movie or brand. Brand owners can pump in more content in support of the brand for positive promotion which may be retweeted by the users to spread positive OWOM.

REFERENCES

[1] Y. Liu, "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," J. Mark., vol. 70, no. 3, pp. 74–89, 2006.
[2] C. M. K. Cheung and M. K. O. Lee, "What drives consumers to spread electronic word of mouth in online consumer-opinion platforms," Decis. Support Syst., vol. 53, no. 1, pp. 218–225, 2012.
[3] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," ACM Comput Surv, vol. 49, no. 2, p. Article 28; 1-41, 2016.
[4] R. V Kozinets, K. de Valck, A. C. Wojnicki, and S. J. . Wilner, "Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities," J. Mark., vol. 74, no. 2, pp. 71–89, 2010.
[5] S. A. Einwiller and S. Steilen, "Handling complaints on social network sites - An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies," Public Relat. Rev., vol. 41, no. 2, pp. 195–204, 2015.
[6] D. Murthy, "Twitter and elections: are tweets, predictive, reactive, or a form of buzz?," Inf. Commun. Soc., vol. 18, no. 7, pp. 816–831, 2015.
[7] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting Elections with Twitter : What 140 Characters Reveal about Political Sentiment," pp. 178–185, 2010.
[8] J. Brown, A. J. Broderick, and N. Lee, "Word of mouth communication within online communities: Conceptualizing the online social network," J. Interact. Mark., vol. 21, no. 3, pp. 2–20, 2007.
[9] I. Taxidou and P. Fischer, "Realtime analysis of information diffusion in social media," Proc. VLDB Endow., vol. 6, no. 12, pp. 1416–1421, 2013.
[10] H. Park and Y. K. Kim, "The role of social network websites in the consumer-brand relationship," J. Retail. Consum. Serv., vol. 21, no. 4, pp. 460–467, 2014.
[11] J. Pfeffer, T. Zorbach, and K. M. Carley, "Understanding online firestorms: Negative word-of-mouth dynamics in social media networks," J. Mark. Commun., vol. 20, no. 1–2, pp. 117–128, 2014.
[12] C. Sarkar and S. Jalan, "Social patterns revealed through random matrix theory," EPL (Europhysics Lett., vol. 108, no. 4, p. 48003, 2014.
[13] R. Solis, "Visualizing stock-mutual fund relationships through social network analysis," Glob. J. Financ. Bank. Issues, vol. 3, no. 3, pp. 8–22, 2009.
[14] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," J. Comput. Sci., vol. 2, no. 1, pp. 1–8, 2011.

[15] R. Vossen, “Does Chatter Matter? Predicting Music Sales with Social Media,” Multicult. Educ. a Plur. Soc., 2011.

[16] S. Moon, P. K. Bergey, and D. Iacobucci, “Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction,” J. Mark., vol. 74, no. 1, pp. 108–121, 2010.

[17] H. Choi and H. Varian, “Predicting the Present with Google Trends,” Econ. Rec., vol. 88, no. SUPPL.1, pp. 2–9, 2012.

[18] E. Kim, Y. Sung, and H. Kang, “Brand followers’ retweeting behavior on Twitter: How brand relationships influence brand electronic word-of-mouth,” Comput. Human Behav., vol. 37, pp. 18–25, 2014.

[19] W. He, S. Zha, and L. Li, “Social media competitive analysis and text mining: A case study in the pizza industry,” Int. J. Inf. Manage., vol. 33, no. 3, pp. 464–472, Jun. 2013.

[20] L. Qiu, H. Rui, and A. Whinston, “Social network-embedded prediction markets: The effects of information acquisition and communication on predictions,” Decis. Support Syst., vol. 55, no. 4, pp. 978–987, 2013.

[21] M. M. Mostafa, “An emotional polarity analysis of consumers’ airline service tweets,” Soc. Netw. Anal. Min., vol. 3, no. 3, pp. 635–649, Sep. 2013.

[22] Y. Wan and Q. Gao, “An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis,” Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015, pp. 1318–1325, 2016.

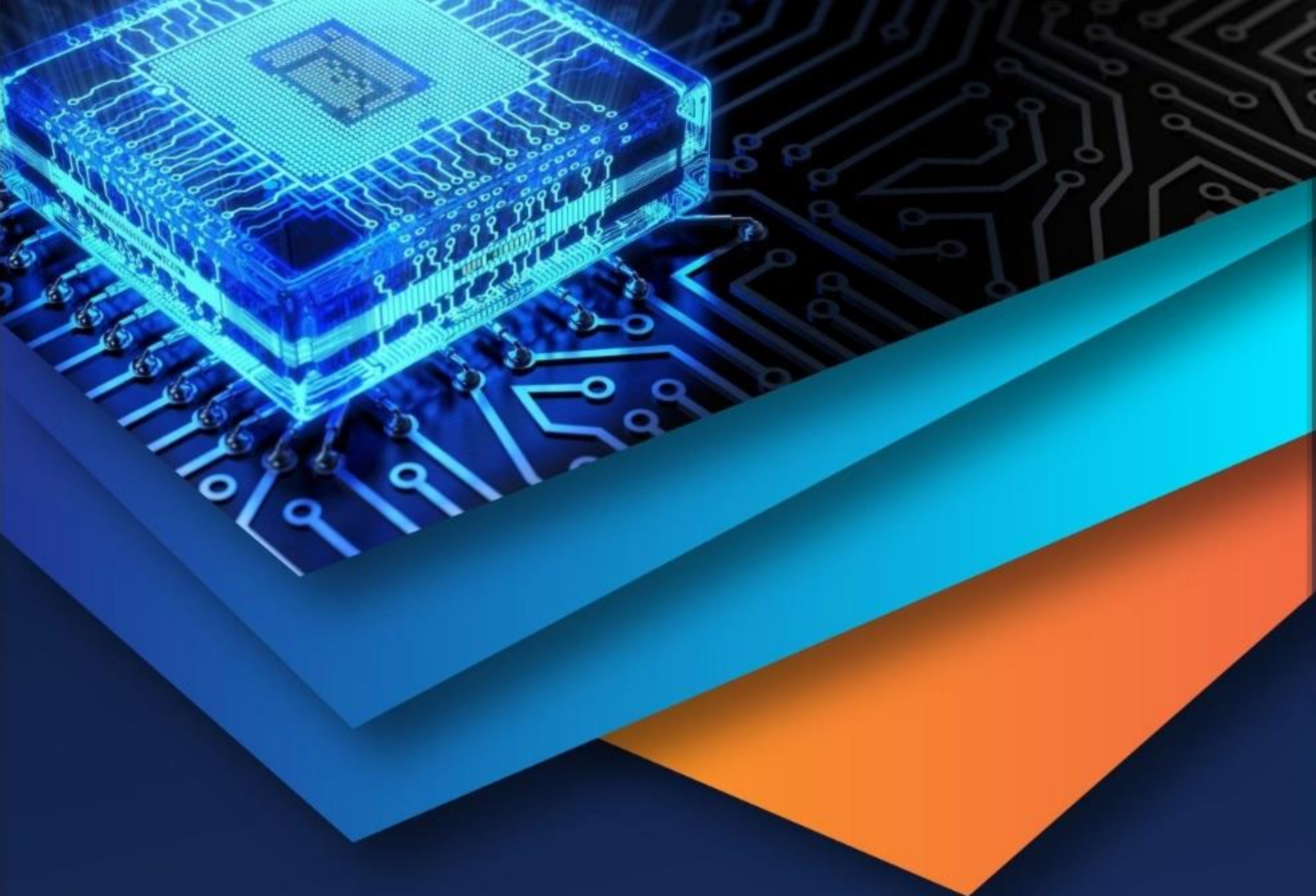
[23] W. Verbeke, D. Martens, and B. Baesens, “Social network analysis for customer churn prediction,” Appl. Soft Comput. J., vol. 14, no. PART C, pp. 431–446, 2014.

[24] B. A. Campbell, “American Economic Association Word-of-Mouth Communication and Percolation in Social Networks Author (s): Arthur Campbell Source : The American Economic Review , Vol . 103 , No . 6 (OCTOBER 2013), pp . 2466-2498 Published by : American Economic Associa,” Am. Econ. Rev., vol. 103, no. 6, pp. 2466–2498, 2016.

[25] J. Feng and P. Papatla, “Advertising: Stimulant or Suppressant of Online Word of Mouth?,” J. Interact. Mark., vol. 25, no. 2, pp. 75–84, 2011.

[26] S. Ranjan and S. Sood, “Identifying network communities in Bollywood Tweet dataset,” in Proceedings- International Conference on advance Studies in Engineering and Sciences, 2017, pp. 503–512.

[27] M. A. Smith, L. Raine, B. Shneiderman, and I. Himelboim, “How we analyzed Twitter social media networks with NodeXL,” Pew Res. Cent., pp. 1–23, 2014.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY