# SOCIAL MEDIA OUTREACH AND SENTIMENT ANALYSIS

Arshpreet Kaur Chawla (M. Tech Scholar) Harkesh Sehrawat (Asst. Professor) Aarushi Jindal (M. Tech Scholar)
UIET, M. D. University UIET, M. D. University UIET, M. D. University
Rohtak (Haryana) Rohtak (Haryana) Rohtak (Haryana)

***Abstract***

***This paper summarizes about analysis on indicators behind requirement for Sentiment Analytics on social media. Sentiment analysis deals with computation of opinions expressed in text. Online networking sites give an abundance of obstinate information which created the requirement for analysing underneath sentiments. Researchers extract data on active social media users and observe patterns of perforation about news/stories among network. The Social network plays a crucial role for information broadcast and that network structure affect extent of information flow. With the advent of web 2.0 social media platforms such as Facebook, Twitter, Instagram, LinkedIn, Interactivity has increased and users have started creating content and moved from earlier pattern of read only usage. Also as the information in rapidly flowing on the web, it is easily accessible to everyone; anywhere; anytime. Even the smallest of activity initiated at a very small scale can be raised globally because of ease and interconnectivity. More often marketers spam on these Medias instead of building genuine relationship. It is widely accepted that twitter is favourite news media, despite being a social network. Therefore we have first highlighted the Social Media penetration & then summarized work done under the hood of Sentiment Analysis on the data scrapped from these social media network sites- particularly Twitter considering special case of Demonetization of Rs 500 & Rs.1000 notes.***

***Keywords: sentiment analysis; twitter; demonetization; social media***

## 1. INTRODUCTION

"Sentiment analysis is the study of people's opinions, attitudes, feelings, and emotions discuss any object such as entities, events, topics, product, issues, services, etc. are respected for extraction of useful subjective information out of the text"[1]. Sentiment Analysis involves the usage of NLP (Natural Language Processing), Statistics, Machine Learning Algorithms to extract, find and depict the sentiment or emotion of a textual matter which is also known as opinion mining[2]. Having large universe of real-word application ranging from Voters preference in coming elections to impact of new Law policy, consumer response on ad campaign to flame detection. This domain has gained a lot of attention, giving researchers new direction for exploration and industry a way to boost their sales, polish brand value in addition to taking decision backed by data to proactively stay ahead with their competitors. Although it is an exciting domain, but there are several challenges involved such as the complexity involved in expressing opinion, sarcasm, detecting analogies, negation, etc [3]. Generally sentiment Analysis is classification of given data points in a data set into different segments wherein the text is given a polarity like Positive, Neutral, and Negative along with its degree. In order to perform Sentiment Analysis we need data, with widespread availability of high speed internet and evolution of Web 2.0 huge volume of unstructured data is generated often called big data.

Social networking sites has gained a lot of popularity and user base in very short span of time hence it plays a vital role in understanding user behaviour. It is a very easy and appropriate way to mobilize thoughts or information through these digital media. Usually customers have a tendency to look for feedback or opinions, and social media network being an open universe gives details on every possible entity. As people have moved towards social network platforms to express their feelings, views and experiences related to product and services, it gives the product or service provider a chance to improve and become more customer oriented by performing appropriate Sentiment Analysis. Sometimes it is also helpful in doing sample study by experimenting or testing new launches on small set of users before rolling out to full audience to attain best results. There are a lot of social media platform namely Facebook, Twitter, Instagram, LinkedIn, etc. After becoming member of any social network, they tend to connect and interact with other user and have a tendency to share data socially which can further be propagated to others. Facebook Statistics [4] show that on average a user has 130 friends in their network, hence if all users in a persons network share some data then that person will receive 130 different posts. Twitter stands second in terms of popularity in social media after Facebook. It is microblogging website launched in 2006 where users post messages that are often termed as tweets in a limit 240 characters increased from 140 characters previously. The metric used to identify the frequency of extent to which any topic is discussed is Social Mention which means "the text inclusion of a monitored keyword in a post on social media platform" [5].

## 2. LITERATURE REVIEW

### i.Quantifying influence of user in Twitter; Huge follower base traction[6]

Directed links in social media depict various things such as intimate friendships, common interests, celebrity gossips or hot news. Information flow is established by these directed links and specify the user's influence on others on social media network. A comparative study on three measures of influence: retweets, indegree and mentions was done. These parameters display the dynamics of user's influence pattern over time and topics which have applications in designing viral marketing on social media.

Here indegree means the number of people who follow a user; retweets is the frequency of forwarded tweet; and mention refers to number of times other users mention a users' name or twitter handle. As per Merriam-Webster dictionary influence is "the power or capacity of causing an effect in indirect or intangible ways." In August 2009, data form 54 million accounts was collected which had 1.7 billion tweets. User having less than 10 tweets in their entire lifespan, invalid screen name were ignored and only 6 million active users were taken into consideration for measuring influence factors. Each user was assigned a value for each influence measure and a rank was assigned. Spearman's rank correlation coefficient was used to measure degree of relationship between two rank set. Value closer to +/-1 means high correlation. There were three findings from the study which are: First, indegree alone has very less relation with influence of a user, popular users having high indegree are not certainly influential in terms of heavy mentions or retweets. Second, most influential users possess significant influence on diverse range of topics. Third, influence is not gained instantly or by surprise, but via consistent effort such as restricting tweets to a solitary point. Considering all three measures, top influencers were recognizable websites and public figures. There was only a marginal overlap of user that topped all three influence measure list.

### ii.Role of Social Networks in Information Diffusion[7]

Information diffusion is a process where new ideas are communicated and that information is shared via some medium over a period between members of that network. This study is focussed on place that social networks possess for information diffusion by running an experiment which randomly displays signs on friends information sharing from 253 million subjects. The persons who are exposed to information have likelihood to share information when compared with ones who are unexposed to information. Additionally, it was also highlighted that strong ties which are less in proportion have more influence in their own cluster or network whereas the number of weak ties are in abundance have major contribution for information propagation. 2 people may propagate similar information as they might get information from same sources such as Websites, Television. The propensity of persons with like traits to ally with another- homophily[8] makes it hard to measure corresponding strong and weak ties and weak ties in information diffusion as people are more alike to those with whom they communicate regularly. Facebook, being most widely accepted social networking platform has over 2.2 billion active user every month as of March 31, 2018 from Facebook newsroom page. American users have 48% of their real life contacts on this site [9] and many of them exchange information on regular basis. A study was done on 7 weeks of data having 253,238,367 subjects, 75,888,466 URLs, and 1,168,633,941 unique subject-URL pairs. Three feasible ways were proposed that explained diffusion of content: (a) user shares any link of Facebook, & subject to this information on news feed causes another user called friend to re-share the same link. (b) Friend also come across same website and share its link on Facebook irrespective of other in the network. (c) A person share link on Facebook and external medium, and exposure with externally shared link cause friend to share the same link on Facebook

### iii.Analysing and predicting news acceptance on twitter[10]

Twitter is integral news media source in online world today and the most important way used for news propagation is retweeting. Supernodes (news medias) having many followers are crucial when it comes to source of information, hence prediction of news popularity within initial few moments of publishing and apprehension of news retweeting propagation from supernodes is highly integral for tasks like online advertisement, monitoring social media campaigns. Based on the features identified from news propagation from suprnodes a news popularity prediction model was build that could analyse and predict final count of retweets of news published instantly. The model has capability to predict (a) total count of retweets from supernode, (b) total retweets in a given hop distance form supernode, (c) number of retweets after a particular time after its published, and (d) count of total retweets upon its saturation. In conclusion it was found that interactions frequency between news source and retweeters has a correlation with popularity of news. News with negative sentiment has some relation with retweet popularity whereas a tweet with positive sentiment has no such proven correlation.

### iv.Sentiment Analysis at User-Level data in` Social Networks[11]

This paper depicts the details on social relationship that can be used to enhance the user-level sentiment analysis, connected users tend to have similar views and opinions and understand how relationship information can complement users' view point according to their expressions. The purpose of this paper, to investigate social network lays foundation to support sentiment analysis, shows an interesting research direction in social network mining. Twitter was used as data source and models were build using semi-supervised user level framework. The model was based on follower/followee network which not only depend between views of a user and the views expressed in his/her tweets, but also between his/her opinion and those of the users that he/she follows. The authors tried some prelim experiments with a Markov Random Field formulation, although the sparsity of the graphs was issue in applying that approach. It focussed mainly on user level instead tweet level as people's thinking process is final outcome for opinion mining.

Their research found a correlation among users sharing same thought with whether they have any connection on social network. Additionally graphical models having social network details resulted in statistically significant enhancement in sentiments categorisation compared to method using only textual details .User-level sentiment analysis was improved by including link information from twitter. Those links correspond to attention, such as when a Twitter user wants to pay attention to another's status updates, or homophily[8]-the idea that similarity and connection tend to co-occur, or "birds of a feather flock together"[12], where person knowing their acquaintances are connected. Choice of follower/followee network vs @ network and directed vs. mutual connections represent different aspects of the homophily vs attention alternatives. It was observed and evident that homophily alone was inferior to collectively accounting homophily and attention. Although some exceptions were also observed. Regardless,

measurable gains can be achieved even when the underlying graph is very sparse, as far as there is a strong dependency among user connectedness and common sentiment. Further it showed many potential future dimension for exploration. A simple job would be to build a larger labelled dataset on more common subjects. Datasets from other online social media systems with other kinds of social networks and more information on users would also be worth exploring. Going forward, several models and semi supervised learning algorithms for exploring network structures should be a worthy try. Finding which parts of the whole network are helpful with respect to prediction on a topic is another interesting direction. Recently, there has been some work on sentiment analysis on Twitter, focusing on the tweet level. The paper concluded by emphasis on need to build a theory on why and how users correlate on different topics in different kinds of networks as an intriguing direction for future research.

### v.Analysis on Sentiment on Citizens of INDIA upon Demonetization using R[13]

Government of India took a major step on 8 November 2016 for demonetization of ₹500 and ₹1000 banknotes, banning the usage of high denomination currency i.e ₹500 and ₹1000 notes as legal tender. This paper introduced a lexicon-based approach to segregate the tweets on #Demonetization in terms of positive, negative and neutral polarity and conclude the response of citizens. The paper focused on how people of India felt post demonetization and thus to present the results of analysis in the shiny dashboard- A web tool created in R to visualize the results. To perform Sentiment Analysis or Opinion mining, 2 set of tweets data was collected Dataset1 from April 2, 2017 to April 8, 2017, consisting of 4500 tweets and from April 9, 2017 to April 15, 2017, a total of 3790 tweets related to #Demonetization were collected, in order to perform sentiment analysis using R programming language. Two different datasets were used to observe study the trend change from first week to second week of how people responded. The paper followed a series of steps necessary to complete sentiment analysis on Demonetization using the tweets and conducting a fair judgment about this step of banning old currency notes. Tweets related of #Demonetization were extracted from the twitter using twitterR package. Tweets consist a sequence of strings, which were split into pieces like keywords, words, symbols, phrases and other elements called tokens in the job of tokenization. To determine the sentiment behind a text the aggregated sum of the score was calculated. Depending on the calculated score the text is been classified as positive, negative and neutral.

Steps involved: (a) Planning and data collection, (b) Data Pre-Processing, (c) Data Probing (The baseline method or "Bag of Words Approach" has 2 dictionaries – one for Positive and other for Negative tagged words. Tokens in tweet after Data preprocessing are assigned a score and total polarity/score of tweet is obtained by aggregating the individual score of token. If score is >0, means tweet has positive sentiment, score=0 indicates neutral tweet, and score<0 shows negative tweet), (d) ShinyDashboard, (e) Sentiment Analysis on 2 sets of tweets. The analysis on Dataset1 of 4500 depicted 34% negative, 26% positive tweets whereas Dataset2 of 3790 tweets had 25% negative tweets & 24% positive tweets. Overall #Demonetization had negative public response.

### vi.Indian Government Demonetization of 500 & 1000 rupee banknotes leading Sentiment Analysis[14]

This paper also takes into consideration the impact of Demonetization policy introduced by Indian Government in order to deal with the issue of black money and perform a Sentiment Analysis and evaluate views or opinion of Indian citizens using Social Networking platform Twitter. As per the notifications issued by RBI (Reserve Bank of India) in 2016, 86% of Indian currency notes were in the denomination of 500 and 1000 notes [15]. The authors not only did nationwide analysis but also performed state-wise analysis. To run the analysis for public behaviour on this governmental policy, Data from Twitter was gathered in 2 different time intervals: 18,926 Tweets during Nov 8, 2016 - Nov 16, 2016 (phase 1) and 11,294 Tweets during Nov 17, 2016 - Nov 23, 2016 (phase 2) in order to conduct an unbiased analysis of what Citizens of India experienced right after the policy was introduced, that is the closure of initial week of note ban, and how that pattern changed in the week following first week of demonetization. Additionally it showed significance of Geographic factor (state) of tweets. Tweetnvi API [16] was used to fetch Tweets in Visual Studio 2012 featuring #Demonetization hashtag. Sentiment Analysis was performed on collected tweets to identify person's feeling using meaningcloud's [17] API that may be integrated as an add-on to Microsoft Excel. This API integrated in MS Excel gives response from following 5 values: P + (high positive), P (positive), Neu (neutral), N (negative), N+ (high negative). This study depicted that large chunk of population seemed satisfied from new policy. In beginning (phase 1) the sentiment was more towards the negative side because the common man had to suffer many hardships. As soon as the new banknotes became readily available (phase 2) & floated in market, the overall sentiment of the people became positive. Looking at more granular level i.e State wise data, 9 amongst 29 States where the opinion was negative had acceptable reason ranging from (a) High rural population, (b) Limited computer & mobiles with internet penetration, (c) Agriculture as main occupation as November was harvesting season. Most of the States had Positive response, Delhi along with 2 other states grouped under Neutral category whereas 7 states were classified with no data.

### vii.Predicting Election with Twitter: What 140 Characters Reveal about Political Sentiment [18]

The authors of this paper took into consideration the German federation election to find out if tweets floating on Twitter reflected the behaviour as seen offline. In their study LIWC (Linguistic Inquiry and Word Count) text analytics software was used on 104,003 tweets messages generated in 4 weeks prior to election which were either related to some Politician or any Political Party. The tweets were originally downloaded in Deutsch language and automatically converted in English so that it could be analysed by LIWC dictionary. They considered two different strands to check if Twitter activity can be used to predict election result: First, contrasting the proportion of recognition a party got with election result; Second, an analysis to check if tweets could tell the correlation of ideas among parties and possible alliance post-election. Key findings from the research depicted that a small number of active, heavy users used Twitter as a medium for Political rumination and number of tweets mentioning a party can be regarded

as a hint for voter choice which stands very near to trivial election polls. 40% of the messages were under the name of only 4% users. In nutshell, it was declared that Twitter can be regarded as a true reflector of political sentiments.

## 3. CONCLUSION

In spite, the fact that the scientists have endeavored numerous sub-problems and presented several arrangements, of which not even single sub-problems have been understood totally. In last few years there has been significant progress in the applications of sentiment analysis. This can be seen from massive start-up culture and some well-known organisations that offer services related to Sentiment Analysis. Industry has a huge demand for these services as all industries and business needs to know consumers perception on their products and services, and also stay at par with their competitors. Earlier people relied on their acquaintances for any type of opinions or advice related to any entity, like which hotel is better in their city, which car should they buy. These days people have switched towards a different way of looking for opinion of others on web before purchasing and referring to any products/services.Various organizations including public, private and government are vesting their resources in analyzing public reviews about their products, brand image, policies and much more. These everyday requirements and the specialized daring tasks on consumer behaviour will retain sentiment analysis energetic and exuberant for a considerable time in future.

It is evident and obvious that areas where Sentiment Analysis needs to be applied will rise and with that standardization of analysis techniques will also come into action for many products. Researchers need to work in a direction so that they seize the meaning of any given sentence in a more descriptive way having threads of interactions and links to details beyond a tweet. The future work in the domain of sentiment analysis systems requires more in depth common sense knowledge warehouse. More relevant understanding needs to be added supporting reasoning principles motivated by persons thinking and backed by logical, psychological and sociological proofs. It will result in-depth, enhanced knowledge of natural language and human thought process and will coherently fill the spaces between machine-processable data or commonly known as structured data and multimodal information or unstructured form of data. Detecting sarcasm still holds a lot of work to be done and some studies could detect it around 56% only [19]. As most of the traffic and data is generated form mobile devices like smartphones or tablets which give access to users geo-location. This geo-location in data set opens a new direction to give useful insights and improve the prediction capabilities to a great accuracy.

## 4. REFERENCES


[1] Al-Asmari, Salma & Dahab, Mohamed. (2017). Sentiment Detection, Recognition and Aspect Identification. International Journal of Computer Applications. 177. 975-8887. 10.5120/ijca2017915675

[2] Zhang L., Liu B. (2017) Sentiment Analysis and Opinion Mining. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA

[3] https://lct-master.org/files/MullenSentimentCourseSlides.pdf

[4] Viswanath, B., Mislove, A., Cha, M. et al: On the evolution of user interaction in Facebook. (2009)

[5] http://www.upwell.us/whats-social-mention/index.html

[6] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: the million follower fallacy. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010) (pp. 10-17). Menlo Park, CA: AAAI Press

[7] Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L.A. (2012). The role of social networks in information diffusion. WWW

[8] P. F. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. In M. Berger, T. Abel, and C. H. Page, editors, *Freedom and Control in Modern Society*, pages 8–66. New York: Van Nostrand, 1954

[9] K. Hampton, L. S. Goulet, L. Rainie, and K. Purcell. Social networking sites and our lives. Technical report, Pew Internet & American Life Project, 2011

[10] B. Wu, H. Shen. Analyzing and predicting news popularity on twitter. International Journal of Information Management 35 (6) (2015) 702–711. https://doi.org/10.1016/j.ijinfomgt.2015.09.004

[11] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P.: User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD (2011)

[12] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27:415– 444, 2001

[13] Goel S. Kumar, Patil S., (2007). Twitter Sentiment Analysis of Demonetization on Citizens of India using R. International Journal of Innovative Research in Computer and Communicative Engineering

[14] P.Singh, et al., Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government, ICT Express (2017), http://dx.doi.org/10.1016/j.icte.2017.03.001

[15] RBI Annual Report, 2016; https://rbidocs.rbi.org.in/rdocs/annualreport/pdfs/0rbiar2016cd93589ec2c4467793892c79fd05555d.pdf

[16] Tweetinvi API, https://www.nuget.org/packages/TweetinviAPI/. Accessed on 10th December, 2016

[17] Meaningcloud API, https://www.meaningcloud.com/products/ excel-addin

[18] Tumasjan, Andranik & Oliver Sprenger, Timm & Sandner, Philipp & Welpe, Isabell. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. Word. Journal Of The International Linguistic Association. 10

[19] B. Liu, Sentiment Analysis and Opinion Mining. CA: Morgan & Claypool Publishers, 2012, ch. 1