A CASE STUDY ON DELHI MCD ELECTION PREDICTION USING SOCIAL MEDIA ANALYTICS

Neetu Narwal
Assistant Prof., Department of Computer Science
Maharaja Surajmal Institute,GGSIP University
Delhi, India.

Kavita Pabreja
Associate Prof., Department of Computer Science
Maharaja Surajmal Institute,GGSIP University
Delhi, India.

Abstract: The growing popularity of social media to share person thought regarding any topic has opened an arena where popularity and sentiment to people regarding various political parties in India can be explored. Social media has witnessed voluminous data related to different political campaigns in the different parts of the world. Researchers have provided analysis for various political campaigns prior to election process and post election as well. In this paper, Twitter has been used as a forum to understand the sentiments of citizens of India towards various political parties viz. BJP, Congress and AAP, during pre-election process. Municipal Corporation of Delhi elections were held on 23rd April 2017 and we have analyzed the tweets originating in India few days prior to election date. The emotions of public in terms of anger, anticipation, disgust, fear, joy, sadness, surprise have been extracted based on their live opinion. These emotions have been related to the elections results and it has been observed that tweets are a valid indicator of political sentiments and it is feasible to use social media to predict electoral results.

Keywords: Twitter, Wordcloud, Frequent terms, associations, Sentiment analysis, BJP, AAP, Congress.

I. INTRODUCTION

In any Democratic Country the Elections are the means of representing the people's choice for electing their Leaders. India being one of biggest democracy, people have profound interest in elections. There always has been curiosity to predict an election outcome. The Electronic Media conducts exit polls to predict the Election outcomes. Lately, it is observed that Exit Polls fail to make an accurate prediction. Furthermore, traditional polls are too costly, and are based on limited amount of data collected after interviewing people. People often hesitate in disclosing the vote cast information in person.

The scientific community has been successful in predicting the outcomes of many complex real life situations. The accuracy of prediction model depends purely on the data. Recent years have witnessed the explosive growth of the usage of social media to share their voice and information using Internet. There are many micro-blogging websites like Twitter, Facebook, and Tumbler etc. Twitter has been widely used Social media tool amongst all age groups of Internet users. Social media generates abundance of data every minute thus provides enormous opportunity to mine the information and get relevant knowledge. These social media data related to politics can be explored to provide some useful outcomes. Lots of successful research work had been done in the area of predicting election outcome based on people voice on social media.

Twitter is one of the widely used social media website amongst users on internet. It allows users to send 280-character messages using their twitter account in the form of short message expressing their thought or following someone's tweet. Tweets related to some political parties or some event may provide valuable information and can be used to know their sentiment towards some event or political party as a whole.

According to Statisca.com website as referred on 15 May 2017 the number of active user in India were approximately 11.5 million in the year 2013 and has shown a tremendous growth of 101.75% and reached 23.2 million Twitter users in the year 2016 [1]. Most recently, the UK general election of 2016 referendum appeared on Twitter's posts.

In this paper we present an analysis of the twitter data for Municipal Corporation of Delhi election, 2017. We gathered Twitter data using the streaming API to extract tweets related to MCD Elections. In the process data is collected in Pre-Poll category.

II. RELATED RESEARCH WORK

Social media has been explored to estimate the popularity of politicians [2], to find out the political interest of social media users [3][4]. Social networking websites are prominently used by entertainment media and political parties to know about people's choice for political preferences [5]. Social media data can be analyzed on hourly basis during an election campaign or party meetings so as to get a detailed insight about the emotions of voters [6]. The real time monitoring and analysis of people communication related to forthcoming election can be compared with actual outcome of the election [7].

There number of researches that used social media data and predicted future election results [8]. In a study author has shown that [9] occurrence of candidate name in the blog post is a good indicator of his/her winning an election poll. In some other study it was observed that the number of facebook supporter is a prominent indicator for success of electoral candidate in forthcoming elections [10][11] . Author [12] in his study used similar notion and used Twitter as social media for predicting the results of forthcoming 2009 German election and indicates that number of tweets are good indicator of vote share.

Author [13] in his study has shown that it is not only the number of tweet but also sentiment of the tweet that provides better indication of people voice for any political party during elections. Author [14] in their work used sentiment classifier

www.manaraa.com

based on lexical induction to provide indication during 2008 presidential election.

In one of the research work done by Author [5] has shown Barack Obama's success in presidential election is correlated with the sentiment expressed by USA Twitter users during that period. In another research work Author [15] has shown the sentiment analysis has predicted the results of Netherlands legislative elections in 2011 and 2012.

III. DATA COLLECTION AND PREPROCESSING

Twitter has been most prominently used for predicting the electoral outcomes in many instances. It has been observed that people comments and their sentiments keep on changing on the basis of electoral promotional events. In our research work we have gathered tweets using twitter API from our account and analyzed more than 3,000 people comments. The tweets published on Twitter's public message board few days prior to the Municipal Corporation of Delhi (MCD) polling on 23rd May 2017, have been collected. We have collected all tweets that contained names of popular political parties contesting for MCD Elections. These parties are Bhartiya Janta Party, Aam Aadmi Party and National Congress.

The people messages on Twitter have relevant information related political parties besides messages have special characters, symbols etc. Thus data collected needs to be cleaned by removing special characters, symbols and stop words from text messages. The word corpus is generated by removing all noise content and converting all text into lowercase characters.

IV. TECHNIQUE APPLIED

In this study we have developed a module in R language to analyze the twitter data related to three major political parties contending for Delhi MCD Elections.

The analysis of Frequent terms and their associations for all three sets of tweets has been accomplished. This has been carried out to understand which terms are more frequently used while discussing about a particular political party.

Cluster of words with hierarchical clustering has been done. Sparse terms are removed so that the plot of clustering will not be crowded with words. After that dendogram showing cluster of words has been generated.

Tweets have also been clustered with the k-means algorithm. Term document matrix of frequent terms has been transposed to obtain a document term matrix. After that kmeans() function is applied to tweets with number of clusters being kept to eight.

We have also done sentiment analysis of these 3000 tweets to understand the emotions of public towards all three mentioned political parties and tried to relate the extracted emotions prior to elections' date with the polling results. We have used many packages viz. syuzhet, lubridate, ggplot2, scales, reshape2, dplyr to facilitate understanding of emotions and plotting the same.

V. RESULTS AND DISCUSSION

All tweets mentioning **"AAP", "BJP", "Congress"** have been analyzed from following perspective:-
- Word cloud generation
- Document matrix of frequent terms
- Cluster dendogram
- K-means clustering

- Sentiment analysis

The Word Cloud has been generated corresponding to tweets mentioning "aap" and it has been found from the size of word that "aap" has been discussed less in comparison of the keyword "mcd elections" as shown in Fig. 1.

Fig. 1 Word cloud "aap"

The Word Cloud has also been generated corresponding to tweets mentioning "bjp" and it has been found from the size of word that "bjp", "mcd" and "elect" has been discussed almost in same count as shown in Fig. 2.
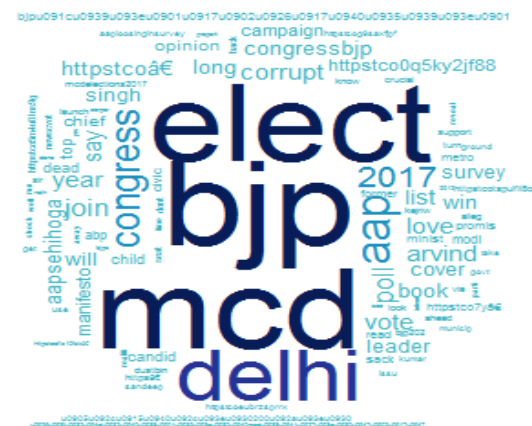

Fig. 2 Word cloud "bjp"

Similarly, the Word Cloud has been generated corresponding to tweets mentioning "congress" and it has been found from the size of word that "congress" has been talked about the most in the tweets. This is shown in Fig. 3.
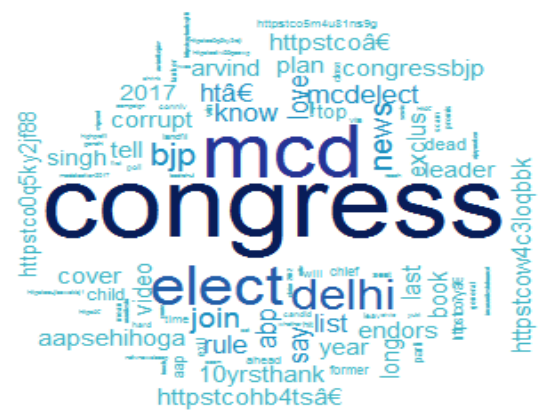

Fig. 3 Word cloud "congress"

www.manaraa.com

The generation of bar plot showing frequent terms with frequency no less than 100 have been done. The most talked words have been plotted as shown in Fig. 4,5,6 for "aap", "bjp" and "congress" tweets respectively. The order of terms by frequency is "mcd", "elections", "aap", "delhi", "kejriwal" and so on in decreasing order for "aap" tweets.

The order of terms by popularity is "bjp", "mcd", "elect", "delhi" and so on in decreasing order for tweets mentioning "bjp". Fot the tweets mentioning "congrees", the order of terms by decreasing order of popularity is "congress", "mcd", "elect", "delhi".
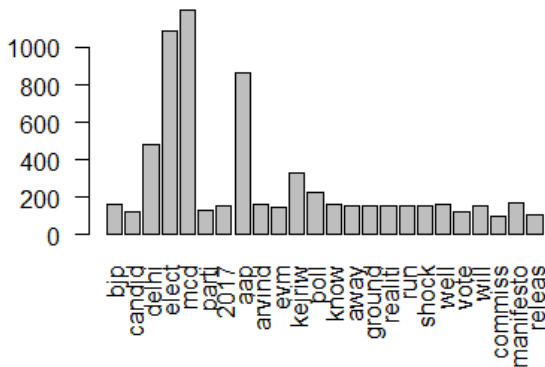
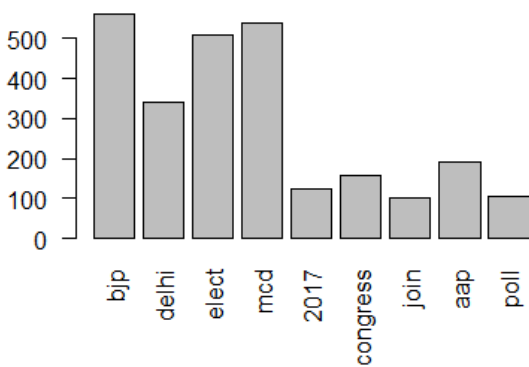Fig.4 Bar plot showing frequent terms from "aap" tweets

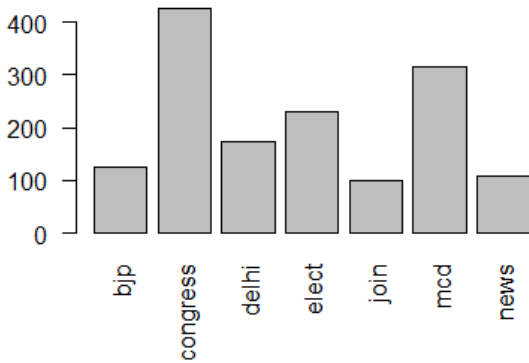Fig.5 Bar plot showing frequent terms from "bjp" tweets

Fig.6 Bar plot showing frequent terms from "congress" tweets

Hierarchical clustering of all three sets of tweets has been done and dendograms have been generated. For each set of tweets k-means clustering technique has also been applied and the words that are talked about together have been extracted.

Corresponding to tweets mentioning "aap" political party, the dendogram depicts that "kejriwal" and "loosing" terms are discussed together as shown in Fig. 7. From clusters, it is clear that when people discuss about "aap", they tend to discuss about "evm" and "delhi" also. The same was observed after the declaration of results when "aap" candidates had put blame on opponent parties about hacking of "evm" machines. The clusters also make it clear that during the list year aap was observed as corrupt. The following eight clusters have been generated.

cluster 1: corrupt list year
cluster 2: vote aap mcd
cluster 3: aap mcd away
cluster 4: parti arvind kejriw
cluster 5: aap evm delhi
cluster 6: aap elect mcd
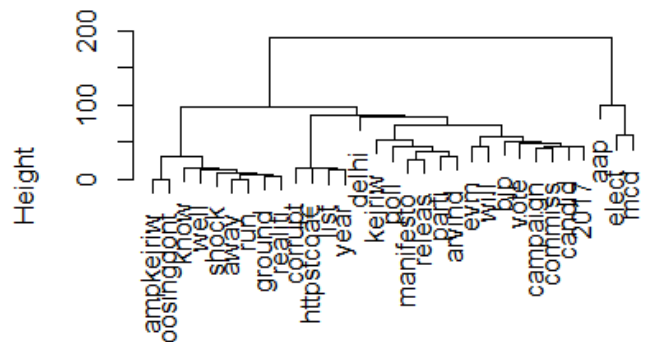cluster 7: mcd poll evm
cluster 8: elect mcd delhi

Fig. 7 Cluster dendogram for "aap" tweets

Corresponding to tweets mentioning "bjp" political party, the dendogram depicts that people have been discussing that "congress" is "dead", as shown in Fig. 8. From clusters, it is surprising to observe that "arvind", "congress" and "join" have been discussed together many number of times publically. The following eight clusters have been generated.

cluster 1: arvind congress join
cluster 2: book congressbjp corrupt
cluster 3: aap bjp candid
cluster 4: leader congress bjp
cluster 5: mcd aap elect
cluster 6: poll mcd opinion
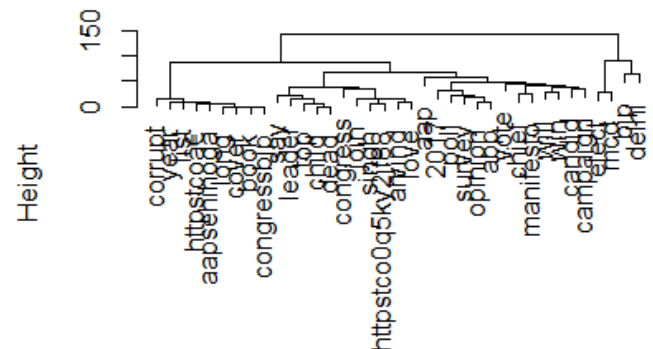cluster 7: mcd elect bjp
cluster 8: bjp delhi chief

Fig. 8 Cluster dendogram for "bjp" tweets

www.manaraa.com

Corresponding to tweets mentioning "congress" political party, the dendogram shows many different clusters but there is no clarity as to what is being talked about as shown in Fig.9. But from clusters generated using k-means clustering algorithm, it is clear that depicts that "leader" would be from "congress" or "bjp" as shown in cluster 3 and cluster 8.The following eight clusters have been generated.

cluster 1: arvind bjp child
cluster 2: 2017 congress elect
cluster 3: congress bjp leader
cluster 4: congress abp exclus
cluster 5: congress mcd elect
cluster 6: delhi arvind congress
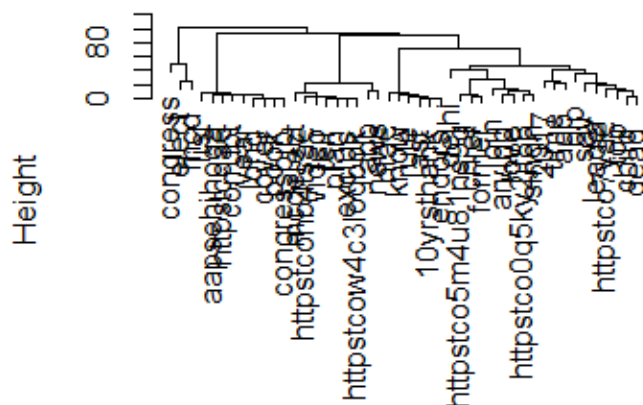cluster 7: mcd congress elect
cluster 8: leader congress bjp

Fig. 9 Cluster dendogram for "congress" tweets

Finally the sentiment analysis of all three sets of tweets has been done to understand the emotions of public prior to day of mcd elections. The get_nrc_sentiment() function from syuzhet package is used to express sentiments of tweets in eight different sentiments anger, anticipation, disgust, fear, joy, sadness and surprise. From the sentiment analysis of "aap" tweets it is very clear that there is enough Anger and anticipation for "aap" party and the same was visible in the results of elections where "aap" suffered a huge loss. As was also apparent from the results that "bjp" had won by great margins and in great numbers, same is clearly observed in sentiments of people. There is too much of Joy and anticipation for "bjp" in the tweets.

As was observed from the results of mcd elections that "congress" was defeated badly, so is the sentiment prior to elections. There is Disgust and anticipation for "congress". The three barplots showing sentiment analysis for "aap", "bjp" and "congress" have been shown in Fig. 10, 11, 12 respectively.

The results of mcd polls are given below in table I that verifies the sentiments of public.

Table I MCD election results

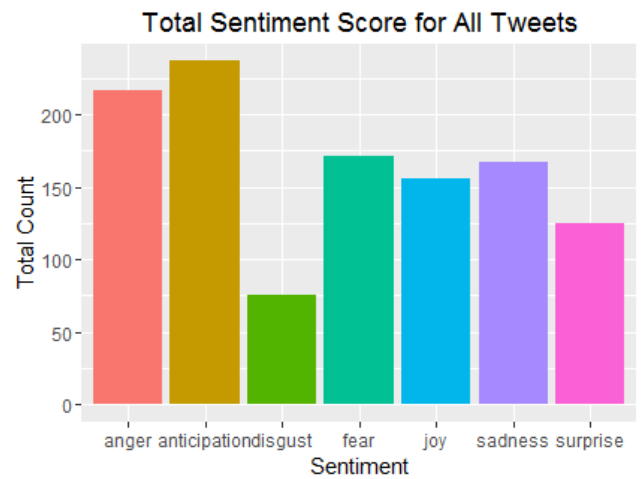| Party name	Number of seats won
AAP	47
BJP	184
Congress	30

Fig. 10 Graph of Sentiment analysis for "aap" party
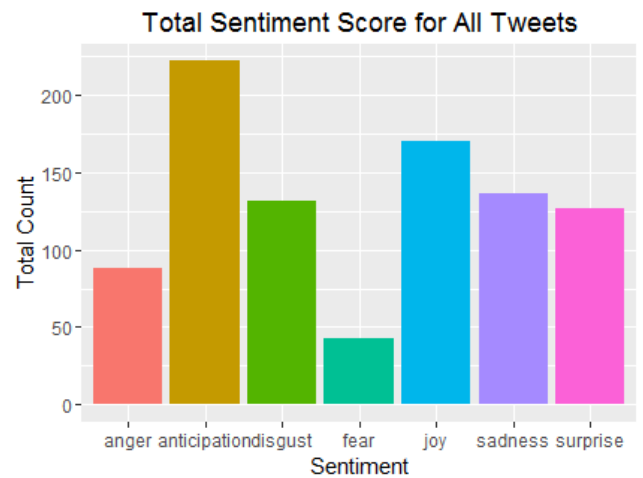

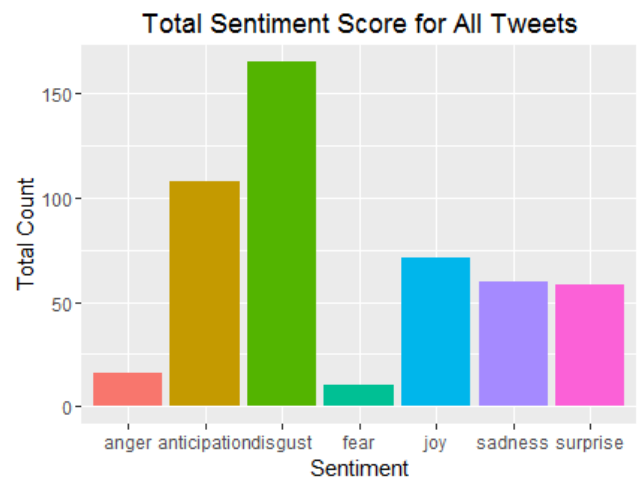Fig. 11 Graph of Sentiment analysis for "bjp" party

Fig. 12. Graph of Sentiment analysis for "congress" party

VI. CONCLUSION AND FUTURE DIRECTIONS

The use of social media by all age groups has opened a new dimensions, where people interest and expression can be gathered using these social media websites. This study has focused on similar exploration where people voice can be collected and analyzed for success of political events and election outcomes.

In our study we have gathered and analyzed over 3,000 twitter messages with words viz. "AAP", "BJP", "Congress" two days prior to MCD elections in New Delhi in April 2017. We made the following observations that social media is

www.manaraa.com

gaining popularity as medium for discussion related to any social event or political deliberations by Indian society. In the years to come we will be witnessing a prominent role being played by social media in electoral campaign and getting insight about the people voice in India.

The discussions on twitter are equivalent to traditional discussions and are capable enough to give a fair idea of standing of contesting political parties. We have generated cluster dendograms and k-means clusters that elaborate about the terms talked together and frequently that show opinion of public towards various political parties. We have also done sentiment analysis of emotions of people which shows there is consistent correlation between social media results and the traditional declared results. Hence we can strongly mention social media population is a representative of population of a nation and social media is capable enough to capture emotions of the people prior to elections and has ability to forecast the electoral results.

In future, we plan to convert this analysis in real time corresponding to tweets arriving on temporal scale. Also we can geographically divide and analyze the tweets according to home town or constituency of candidates. We can also do deeper analysis based on names of different contestants. The American Mechanical Turks (AMT) approach can also be made use of so as to eliminate any sarcasm which could not be picked by functions of "R" package and comparison can be done between AMT analysis and analysis by "R" software.

VII. REFERENCES

[1] https://www.statista.com/statistics/381832/twitter-users-india/

[2] Gloor, P. A., Krauss, J., Nann, S., Fischbach, K., & Schoder, D. (2009). Web science 2.0: Identifying trends through semantic social network analysis, Proceedings of International Conference on Computational Science and Engineering, Vancouver,Vol 4, pp. 215–222, 2009.

[3] Barbera´, P. (2012). Birds of the same feather tweet together. Bayesian ideal point estimation using twitter data. Political Analysis 23.1, pp. 76-91, 2014.

[4] Conover, Michael D., Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. "Predicting the political alignment of twitter users." In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, pp. 192-199. IEEE, 2011.

[5] O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. Icwsm, 11(122-129), 1-2.

[6] Jin, Xin, Andrew Gallagher, Liangliang Cao, Jiebo Luo, and Jiawei Han. "The wisdom of social multimedia: using flickr for prediction and forecast." In Proceedings of the 18th ACM international conference on Multimedia, pp. 1235-1244. ACM, 2010.

[7] Perea, Eva Anduiza, Michael James Jensen, and Laia Jorba, eds. Digital media and political engagement worldwide: A comparative study. Cambridge University Press, 2012.

[8] Sang, Erik Tjong Kim, and Johan Bos. "Predicting the 2011 dutch senate election results with twitter." In Proceedings of the workshop on semantic analysis in social media, pp. 53-60. Association for Computational Linguistics, 2012.

[9] Véronis, Jean. "Citations dans la presse et résultats du premier tour de la présidentielle 2007." Retrieved December 15 (2007): 2009.

[10] Upton Jr, G. "Does attractiveness of candidates affect election outcomes." URL: http://com/lib/files/AttractivePoliticians. pdf Google Scholar (2010).

[11] Williams, Christine, and Girish Gulati. "What is a social network worth? Facebook and vote share in the 2008 presidential primaries." American Political Science Association, 2008.

[12] Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. Icwsm, 10(1), pp.178-185.

[13] Chung, Jessica Elan, and Eni Mustafaraj. "Can collective sentiment expressed on twitter predict political elections?." In AAAI, vol. 11, pp. 1770-1771. 2011.

[14] Lindsay, R. "Predicting polls with Lexicon." Available at: languagewrong. tumblr. com/post/55722687/predicting-polls-with-lexicon (2008).

[15] Sanders, E. P., and A. P. J. van den Bosch. "Relating political party mentions on Twitter with polls and election results." (2013).