I.J. Intelligent Systems and Applications, **2017, 12, 20-28**
Published Online December 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2017.12.03

Towards Prediction of Election Outcomes Using Social Media

Vinay K. Jain
Department of CSE, Jaypee University of Engineering and Technology, Guna, 473236, India
E-mail: vinay2588@gmail.com

Shishir Kumar
Department of CSE, Jaypee University of Engineering and Technology, Guna, 473236, India
E-mail: dr.shishir@yahoo.com

Received: 04 March 2017; Accepted: 08 June 2017; Published: 08 December 2017

***Abstract*—**Exploiting social media data by extracting key information from it is one of the great challenges in data mining and knowledge discovery. Every election campaign has an online presence of voters which uses these social media platform to express their sentiments and opinions towards political parties, leaders and important topics. This paper presents a novel data collection technique for prediction of election outcomes and a topic modeling method for extracting topics. Data collection technique used RSS (Rich Site Summary) feeds of news articles and trending keywords from Twitter simultaneously and constructed an intelligent prediction model based primarily on the volume of tweets and sentiment of users. This paper effort to improve electoral predictions using social media data based dynamic keyword methodology.

Different techniques for electoral prediction based on social media data has been investigated based on existing literature and isolate the factors which improve our methodology. Meaningful inferences such as the popularity of leaders and parties during different intervals, trending issues, and important factors are extracted from the data set. The election outcomes are compared with traditional methods used by survey agencies for exit polls and validation of results showed that social media data can predict with better accuracy. The research has identified that data collection technique and timing play an important role in yielding better accuracy in predicting outcomes and extracting meaningful inferences.

***Index Terms*—**Sentiment analysis, Opinion mining, Twitter, Text classification, Feature Selection, Latent Dirichlet allocation(LDA).

I. INTRODUCTION

Social media data have been a primary focus in the field of information retrieval (IR) and text mining due to an excessive amount of unstructured data in real time. Every tweet, comment and blog post might reflect their sentiments. These unstructured data provides valuable knowledge which constitutes a big opportunity for creating new services for governments, businesses or individuals [1]. Exploiting these unstructured data created a new field called opinion mining and sentiment analysis.

In the area of opinion mining and sentiment analysis, which deals with the computational calculation of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions More recently, authors developed sentiment analysis methods that can be used across multiple domains like movie and product reviews, election result prediction; disease outbreak, stock market etc [2].This paper focused on election result prediction using the social media data. Indian social media users have rapidly evolved over the past few years to form a complete ecosystem which deals in several areas such as news, politics, health, government policies and finance [3-5].

Nowadays, every election campaign has an online presence and users use these social media to express their sentiments and opinions towards political parties, leaders and important topics during election [6]. Initial studies often presented optimistic results regarding the predictive capacity of Twitter data relative to election results Authors presented optimistic results regarding the predictive capacity of social media towards the election results in different countries [7].The most important factor for better prediction depends on the data or data collection methodology [8]. If the data collected is not much relevant towards the event then the results may be inappropriate [9]. The proposed research focused on getting relevant data from these social media and develops a prediction model which helps in better understanding of election outcomes and meaningful factors. An intelligent data collection technique has been presented using RSS (Rich Site Summary) feeds from top news agencies of India and trending topics from Twitter. Proposed technique is applied to investigate and illustrate meaningful inferences for Delhi Assembly Election 2015[10].Comparison of various machines learning algorithm for classification of data set is carried out. Prediction results are compared with various survey agencies exit polls for validating the results and found that results based on social media data can predict with better accuracy. Analysis of social media data also helps in decision making such as finding key political topics, the popularity of the candidate, the impact of debate etc.

The rest of this paper is organized as follows: First, evaluation of various techniques used for prediction the election outcomes in the different geographic region with their data collection methodology. Second, present proposed intelligent techniques for data collection. Third, introduce the features that are relevant and correlated with the election outcomes and propose our prediction model. Fourth, prediction results will be presented with a comparison to exits poll results. Finally, draw conclusions from our experiments and propose our future research.

Table 1. Contribution of different authors for election prediction

Year	Authors	Country	Election Type	Method(s)
2010	O'Connor [19]	US	Presidential	Sentiment Analysis using word frequencies
2010	Tumasjan[20]	Germany	Federal	Count Tweets/Hashtags
2010	Hopkins & King[21]	US, France , Italy	Presidential	Sentiment classifier based on lexical induction
2010	Diakopoulos & Shamma[22]	USA	Presidential	Demonstrate visuals and metrics of tweets
2011	Choy[23]	Singapore	Presidential	Reweighting techniques , Count Tweets & Sentiment Analysis
2011	Gayo-Avello[24]	USA	Senate	Count Tweets & Sentiment Analysis
2011	Bermingham[25]	Ireland	General	Count Tweets & Sentiment Analysis
2011	Metaxas et al. [26]	USA	General	Multiple methods
2011	Conover et.al. [27]	USA	Congressional midterm	Network clustering algorithm
2011	Maynard & Funk[28]	UK	Pre-election	NLP techniques
2012	Sang [29]	Dutch	Senate	Count Tweets
2012	Larsson [30]	Sweden	General	Identifying user types
2012	Choy[31]	USA	Presidential	Count Tweets
2012	Jungherr [17]	Germany	Federal	Count Hashtags & Sentiment Analysis
2012	Skoric et al.[32]	Singapore	General	Volume of tweet
2012	Tjong & Bos[33]	Dutch	Senate	Sentiment Analysis
2012	Marquez et al.[34]	USA.	Presidential	time series obtained from Twitter messages
2012	Soler et.al [35]	Spain	Regional and General	Volume of tweets
2012	Wang et al.[36]	USA	Presidential	Sentiment Analysis
2013	Contractor [37]	USA	Presidential	NLP techniques
2013	Cameron[38]	New Zealand	General	Number of Friends/Followers
2013	Nooralahzadeh[18]	US & French	Presidential	Sentiment Analysis
2013	Ceron[39]	Italy & France	Presidential	Count Tweets & Sentiment Analysis
2013	Gaurav[40]	Venezuela	Presidential	Count Tweets & User
2013	Wong[41]	USA	Presidential	Count Tweets/Retweets & Sentiment Analysis
2013	Sanders[42]	Netherlands	Parliament	Count Tweets
2013	Fink[43]	Nigeria	Presidential	Count Tweets & Sentiment Analysis
2013	Kim et al.[44]	USA	Presidential election	Iterative topic modeling algorithm
2013	Bakliwal et al.[45]	Ireland	General	3-class sentiment classification
2013	Then & Ghanem[46]	UK	Members of Parliament	automatic identification
2013	Vaccari et al.[47]	Italian	General election	Volume of tweets
2013	Vergeer & Hermans[48]	Netherlands	general	Volume of tweets
2013	Wong et al.[49]	USA	Presidential	Tweeting and re-tweeting behavior
2014	Song et al[50]	Korea	Presidential	Text-mining techniques
2014	Makazhanov[51]	Canada	General	Count of Interactions between Followers
2014	Ifrim et.al [52]	USA	Presidential	Hierarchical Clustering of Tweets
2014	Mehndiratta et. al.[53]	India	General	Opinion mining
2014	Barbera & Rivero[54]	Spain &US	Legislative and Presidential	Volume of tweets
2015	Vaccari et al.[55]	Italian	General	Volume of tweets
2016	Tunggawan & Soelistio[56]	USA	Presidential	Naive Bayesian predictive model
2017	Jain & Kumar[5]	India	Legislative Assembly	Emotions based analysis

II. Related Work

The growth of social media data has interested researchers from various disciplines to uncover the hidden knowledge by applying intelligent data analysis techniques. There are two main types of textual information on the web: facts and opinions. On one hand, facts are assumed to be true and on the other, opinions express subjective information about a certain entity or topic. Some argue that the usage of Twitter and other social media was one of the reasons Obama won the US presidential election in 2008 [11]. Authors present reasons for the increased success of microblogs such as Twitter, noting that their successes come as a result of their unique communication characteristics [12]. A platform such as Xbox gaming is also used for forecasting election [13].Grosse et al.[14] developed a framework which allows mining opinions from Twitter based on incrementally generated queries. Robertson et al., [15] introduced "realized public sphere" in virtual spaces in contrast to political discourse. Hughes and Palen [16] argue in favor of using microblogs as a public information channel used by authorities, for instance in emergency situations. Jungherr et al., [17] argue that methods of prediction using social media analytics are frequently contingent on a somewhat arbitrary experimental variable. Nooralahzadeh et al., [18] used complete keywords by adding the campaign and election hashtags. Most of the authors focused on the volume of the tweets and sentiment analysis based techniques for predicting the election outcomes. Authors used NLP based techniques for an understanding of tweets [5]. Authors also discussed suitable time period for data collection, which is an important feature. In this paper, data collection started before the election dates declaration from the election commission of India.

Table 1 presents a brief review of electoral prediction using social media with a wide variety of techniques used for mining hidden knowledge from the data sets.

III. Proposed Data Collection Methodology

In this section, proposed intelligent technique for data collection is presented. The important variable for data collection from social media data are the keywords which help in getting relevant tweets [57, 58]. Most research for keyword selection used candidate/party names, election-related hashtags, and campaign hashtags[18].The period for the conduction of election takes almost 2 to 6 months. During this period lots of topics or events related to the election are discussed in social media, which might be important for predicting the outcomes. Taking this assumption that using more keywords related to the election can collect more relevant data and in turn, increase the prediction accuracy. Proposed data collection methodology presented in Fig. 1, is different from other author's techniques; here we considered those keywords which are trending and dynamic. For getting dynamic keywords three types of data sources are used. Firstly, analysis top newspapers RSS feeds and find out top topics using the Topic modeling technique, LDA (Latent Dirichlet allocation) given by [59]. From these topics, filter out most occurred topics and keywords used in data collection. This methodology gives dynamic keywords which are trending during the election and related to public sentiments. Secondly, use of top trending hashtag taken from Twitter and thirdly, collected from tweets.

A. Keyword Selection

In this section, identification of dynamic keywords forms three different methods is presented in Fig. 1. These keywords are used for fetching relevant tweets from Twitter. These methods are:

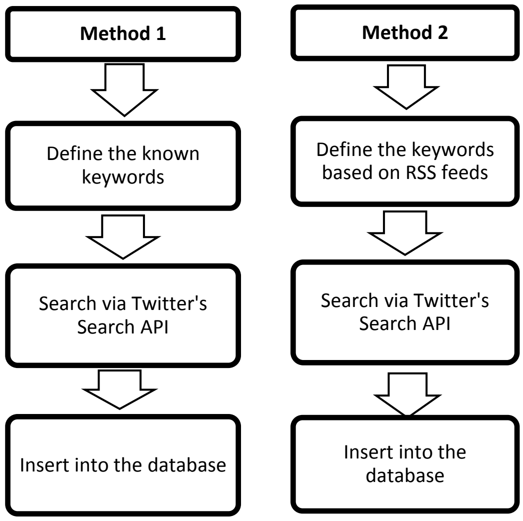

Fig.1. Data collection steps

(i). Keywords from tweets

Keywords selection is based on TF-IDF (Term Frequency–Inverse Document Frequency) approach which measures the importance of words (or "terms") in a document based on how frequently they appear across multiple documents. Weight is composed of two terms: the first computes the normalized Term Frequency (TF) and the second is to find Inverse Document Frequency (IDF) given in Equation (1) and Equation (2).Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Do not use abbreviations in the title unless they are unavoidable.

$$\text{TF(t)}= \frac{[\text{Number of times term t}'\text{appears in a document}]}{[\text{Total number of terms in the document}]} \quad (1)$$

$$\text{IDF(t)} = \log_e \left(\frac{\text{Total number of documents}}{\text{Number of documents with term t in it}}\right) \quad (2)$$

(ii). Keywords from RSS feeds

In this section, RSS (Rich Site Summary) feeds from top news agencies in India have been collected continuously during the period of election and store in the database. Filtering the most important topics are carried out using LDA (Latent Dirichlet allocation) based topic modeling approach. After finding the relevant topics, keywords are filtered out based on ranking. According to Blei et al., [59] model, a number of topics k has to be fixed apriori. For a document w = (w_1,w_2 . . . , w_N) of a corpus D (collection of RSS feeds) containing N words from a vocabulary consisting of V different terms, wi ∈ {1, . . . , V } for all i = 1, . . . , N. The steps are illustrated in Fig. 2.

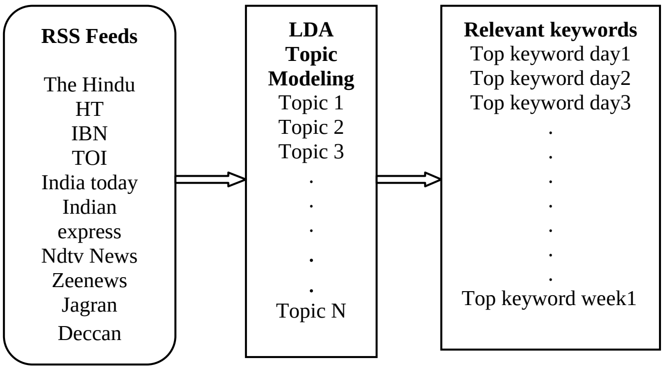

Fig.2. Selection procedure for topics and keywords from RSS feeds

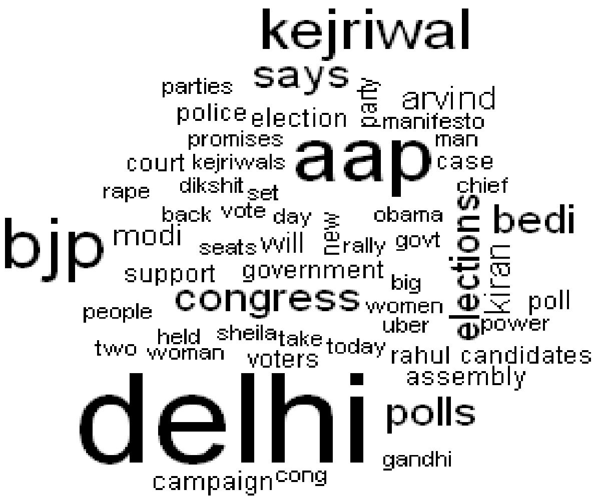

Fig.3. Tag cloud showing most relevant keywords in RSS feeds (02-Jan 15 to 05-Feb 15)

(iii) Trending hashtags from Twitter

In this section, continuously evaluation and collection of top trending hashtag have been performed during the whole election period. There are various similar keywords are found using these techniques during different intervals and shown in Table 1.

Table 2. Tweets collected during different time intervals with top hashtags

Time Period	No. of tweets	Top Keywords
20-Nov-2014 to 20-Dec-2014	166380	#delhielections, #delhipolls, #Kejriwal, #Arvind, ,#CrazyKejaria,#aap, #kejriwal,
20-Dec-2014 to 20-Jan-2014	212448	#CleanPolitics, #vote4aap ,#BJPdelhi, #AAPKiDilli
20-Jan-2015 to 06-Feb-2015	312694	#5saalkejriwal, kejriwal4delhi #abkibaarbedisarkaar,#aapkama nifesto

B. Data Collection

After selection of keywords, data collection process is started by retrieving tweets using Twitter. It offers two APIs: Streaming API and REST API for retrieving tweets. The data collection process started before the announcement of the final dates for the election by the Election Commission of India. Continuous data has been fetched from Twitter started from 20th November 2014 to 07th February 2015 using dynamic keywords given in Fig.3. The collected dataset contains 1085721 tweets of which 390425 are re-tweets or duplicates, leaving a total of 703521 tweets. The dataset is divided into three parts based on different time periods.

C. Data Preprocessing

Every word contains in a tweet is important in decision making, so pre-processing of these tweets is an important task because these messages are full of slang, misspellings, and words from other languages[60-63]. In order to tackle the problems with the noise in texts, normalization of tweets is performed by applying text preprocessing steps like tokenization, stop words removal, stemming, lemmatization, feature weighting, dimensionality reduction, frequency based methods proposed by Bao et .al[62].

IV. DATA ANALYSIS

In this section, examination of tweets with different parameters based on the volume of tweets and sentiment analysis of tweets is examined and various meaningful inferences had been extracted. The data analysis technique provides various meaningful inference which helps in the prediction of election outcomes.

A. Volume based analysis

For detecting the user's voting intention from their tweets towards Delhi Assembly election 2015 has been examined before the declaration of results from Election

Commission of India. All the major political parties such as Party 1(BJP) Party 2(AAP) and Party 3(Congress) are gearing up for the elections and have formed various strategies for effective campaigning in the city.

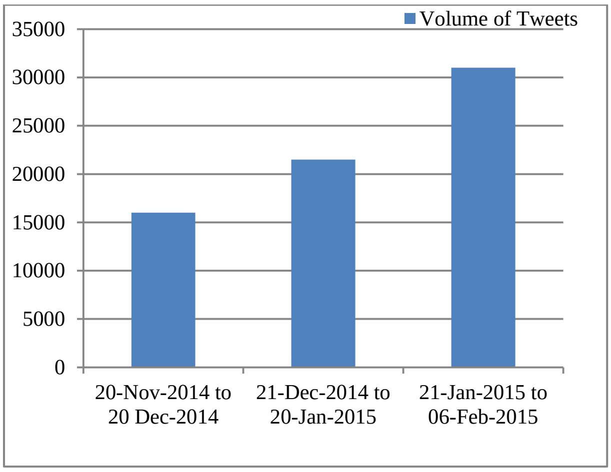

Fig.4. Volume of tweets

The elections are going to be largely between Party 1 and Party 2. The election commission announced the schedule for the elections on 12 January 2015.Date of the poll was 7th February 2015 and counting of the vote is held on 10th February 2015[10].Table 3 and Table 4 presented the total no. of tweets collected from Twitter for corresponding Political parties and their CM candidates.Fig.4 presents total no. of tweets collected.

Table 3. Total tweets count for political party

Total Tweets Counts: 420705			
Political Party	Party 1 (AAP)	Party 2 (BJP)	Party 3 (Congress)
Total	211500	156188	53017

Table 4. Total tweets counts for CM candidate

Total Tweets Counts: 216164			
CM Candidate	Candidate 1 (Arvind Kejriwal)	Candidate 2 (Kiran Bedi)	Candidate 3 (Ajay Maken)
Total	168183	44426	3555

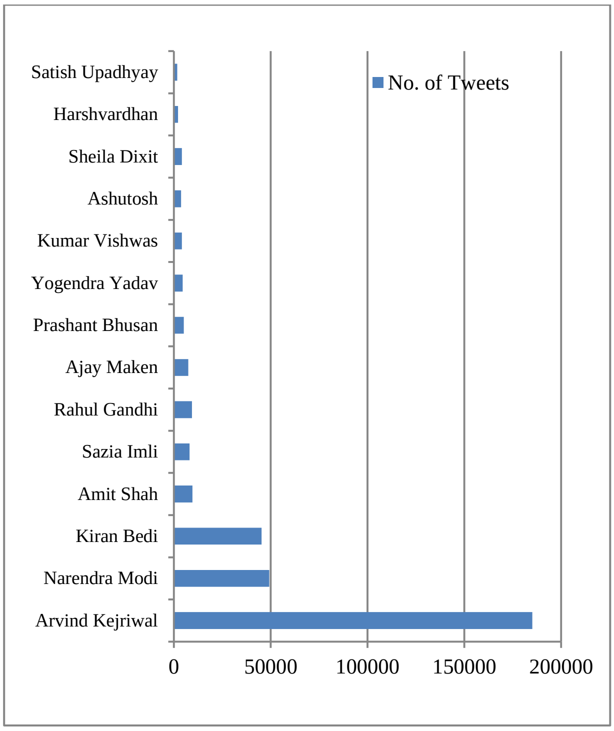

Fig.5. Popularity of leaders during election based on tweets volume

During the election time, we find out the popularity of different leaders and know how much user influenced from them. During the election there are many leaders of different party delivered the speech in different rallies related to Delhi election and also talk to media news channels. The political party leader's influence the voters and how much they are popular during this time interval are shown in Fig.5.Gayo-Avello [24] claimed that "conversation" about elections grew as the campaign progressed and elections came nearer. To verify this claim, extracting more keywords and topics related to the election using LDA-based approach in the dataset produce following results presented in Fig. 6.

B. *Sentiment based analysis*

Proposed framework for sentiment analysis is based on simple opinion and comparison based opinions. Generally, user tweets contain two types of opinions (a) Simple opinion like a vote for Party or vote for Candidate or (b) Comparative opinions like Candidate 1 loyalty vs Candidate 2 or Party 1 vs. Party 2. The system used simple opinion based approach by taking relevant feature for solving the problem of classifying the political sentiment of tweets.

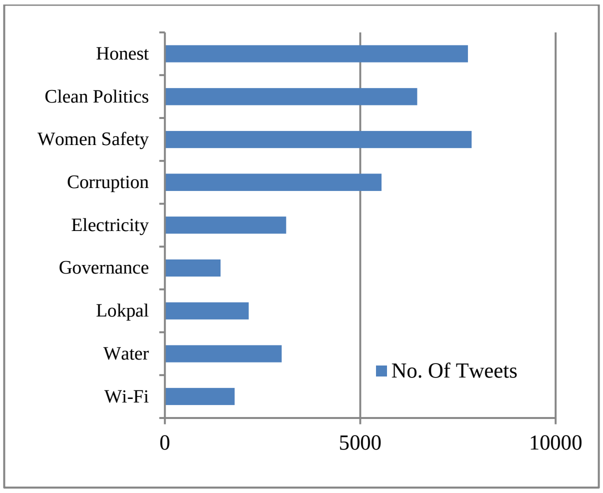

Fig.6. Finding important issues during Delhi election 2015

For building the training and testing set, 1000 tweets during different time intervals are taken. Keywords from the preprocessed dataset are taken as features such as party name, candidate name and their alias. Tweets containing more than one party or candidate name considered as neutral.

For determining the polarity of tweet 4-cross validation of popular machine learning methods are applied such as Support Vector Machine using linear kernel function, Naive Bayes Classifier, Decision tree, Random Forest. Suitable measures are used to assess the efficiency of algorithms, attributes used for classification, training set. Sentiment analysis can help in the analysis the mood for the user.

A comparative performance evaluation of each machine learning methods in terms of the correctly predicted polarity of tweets is examined. The results are explained in terms of precision, recall, accuracy, and F-measure and shown in Table 5. The F-measure provides the overall performance of a classifier and is calculated using the following formula:

$$\text{F} - \text{measure} = \frac{2(\text{Precision})[\text{Recall}]}{\text{Precision}+\text{Recall}} \tag{3}$$

Table 5. Performance of the results obtained for each labeled dataset.

Algorithm	Accuracy	Precision	Recall	F-measure
SVM	79.4%	0.798	0.785	0.792
Naïve Bayes	69.25%	0.685	0.714	0.699
Random Forest	77.25%	0.786	0.749	0.767
Decision Tree	70.80%	0.670	0.74	0.706

From the evaluation, we found that SVM techniques give better results as compared to others machine learning techniques with the accuracy of 79.4%. We get an overall positive sentiment towards Party 1(AAP) and their CM candidate, Candidate 1.

V. Prediction Model

In this section, a prediction model has been developed to know the user political preference toward a party or candidate in term of percentage of votes. For a given user, U= $\{u_1,u_2,u_3.........u_n\}$ and a set of parties P $=\{p_1,p_2,p_3\}$ or Candidate C=$\{c_1,c_2,c_3\}$,examination is carried out as whom u_i has interactions or likely to vote in the upcoming election.

To address the problem, volume based approach has been considered in which a user mentioned the party name, party alias names or the candidate name in their tweets will be considered as the supporter. Using mentioned technique, the percentage of votes for a particular party and the no. of seats it will get has been calculated. Outcomes of the applied technique are presented in comparison to other survey agencies in Table 6 and Fig 7. Results are published before the election results (10-Feb-2015) in Twitter, Facebook and Delhi election forecast (2015) on 03-Feb-2015.

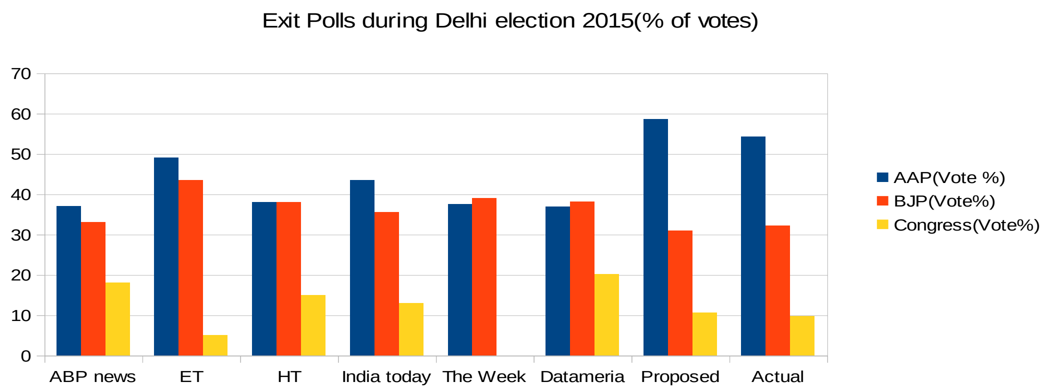

Fig.7.Comparison of exit polls with proposed method [% of votes]

Table 6. Comparison of prediction results of survey agencies with proposed method

Date	Survey agency	Sample Size	Political party		
			Party 1	Party 2	Party 3
5–7 Nov-14	ABP News-Nielsen	6528	28	36	5
12-Nov-14	NewsX–CVoter	2447	26	37	6
21 Nov-5 Dec -14	Economic Times–TNS	7113	22–25	43–47	0–3
4–8 Dec- 2014	ABP News–Nielsen	6409	17	45	7
18-Dec-14	India Today–CICERO Wave1	4273	28±3	37±3	4±1
25-Dec-14	TV24 News India	8200	39	23	5
12-Jan-15	India Today–CICERO Wave 2	4459	25–30	34–40	3–5
12-Jan-15	India TV-CVoter	4238	29	35	5
15-Jan-15	ABP News–Nielsen	6414	28	34	8
11–15 Jan15	News Nation	3195	33±2	31±2	5±1
10–16 Jan-15	Zee News-Taleem	4200	29	37	4
18–24 Jan-15	India TV-CVoter	1306	28	37	5
22–24 Jan-15	The Week-IMRB	4055	29±2	36±2	4±1
24–25 Jan-15	ABP News–Nielsen	6396	35	29	6
10-19 Jan-15	Hindustan Times-C fore	7147	31–36	31–36	2–7
25–31 Jan-15	India TV-CVoter	10862	31	36	2
25–31 Jan-15	Economic Times-TNS	3260	38±2	30±2	2±1
27 Jan- 1 Feb-15	Hindustan Times-C fore	3578	36-41	27-32	7
31 Jan - 1 Feb-15	AAP [internal]	3188	51±6	15±5	4±2
3-Feb-15	NDTV Poll of Opinion Polls	-	37	29	4
3-Feb-15	India Today Group-Cicero	3972	42±4	22±3	5±2
20 Nov-14 -06 Feb-15	Proposed Method	636869	42	21	7

VI. Conclusion

In this paper, a novel intelligent prediction technique for election outcome prediction based on dynamic keywords and topic modeling has been proposed. This technique is applied to electoral prediction during different time intervals using social media data to get the outcomes. This technique is based on the sentiment of voters and volume based technique for prediction. Political Election prediction, results were published on social media, before the declaration by Election Commission of India.It shows how social media sources can help in forecasting election results before the actual results. A brief analysis of important issues regarding elections is also examined. This justifications mentioned in this paper verifies that proposed dynamic keyword based approach can use as next-generation forecasting system for event prediction with better accuracy.

Possibilities of improvement in results are also there, when proposed approach will be applied to the events of those countries where social media is used by most of the citizens for expressing their opinions. Multi-language data preprocessing can improve the prediction accuracy.

References

[1] Vinay K Jain, Shishir Kumar, "Big Data Analytic Using Cloud Computing", in 2nd *IEEE International Conference on Advances in Computing and Communication Engineering (ICACCE 2015)*, Dehradun, pp.667-672, 2015.

[2] B. Pang and L. Lee, “Opinion mining and sentiment analysis”, *Foundations and Trends in Information Retrieval*, vol.2, pp.1–135, 2008.

[3] Vinay K Jain, Shishir Kumar, “An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter”, *Procedia Computer Science*, Vol.70, pp.801-807, 2015.

[4] Vinay K Jain, Shishir Kumar, “A Novel Approach to Track Public Emotions Related to Epidemics In Multilingual Data”, in *2nd International Conference and Youth School Information Technology and Nanotechnology (ITNT 2016)*,Russia,pp.883-889,May 2016.

[5] Vinay K Jain, Shishir Kumar, "Extraction of Emotions from Multilingual Text Using Intelligent Text Processing and Computational Linguistics" *Journal of Computational Science*, 2017.(doi: 10.1016/j.jocs.2017.01.010).

[6] M.Anjaria,, R.M.Guddeti, “Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning”, in *Proceeding of Communication Systems and Networks [COMSNETS]*, 2014 Sixth International Conference ,Banglore, India,pp.1-6,2014.

[7] M.Arias ., A. Arratia, and R. Xuriguera “Forecasting with Twitter Data” *ACM Transactions on Intelligent Systems and Technology*, Vol. 5, Article no.8,2012.

[8] N.Li,,D. Wu,“Using text mining and sentiment analysis for online forums hotspot detection and forecast”, *Decision Support Systems*, vol.48, pp.354–368,2010.

[9] H.Wang, C.Dogan,A. Kazemzadeh,F. Bar & S.Narayanan, “A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle”, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Republic of Korea, pp.115–120,2012.

[10] Delhi Assembly Election, “Delhi Legislative Assembly election 2015”, available at: http://en. wikipedia org /wiki /Delhi_Legislative_Assembly_election,_2015[accessed 15-March-2016].

[11] F.Matthew and S.Dutta "Obama's win means future elections must be fought online", available at: http://www.theguardian.com/technology/2008/nov/07/barackobama-uselections,2008.[accessed 23-Feb-2015]

[12] F.Wigand,L. Dianne, " Tweets and retweets: Twitter takes wing in government", *Information Polity*, vol.16, pp.215-224,2011.

[13] W.Wang,D. Rothschild ,S.Goel,A. Gelmana" Forecasting elections with non-representative polls", *International Journal of Forecasting*, vol.31,pp.980-991,2015.

[14] K.Grosse, M.P. González, C.I. Chesñevar,A.G. Maguitman, A.G., "Integrating argumentation and sentiment analysis for mining opinions from Twitter, *AI Communications*, Vol.28, pp.387-401,2015.

[15] S.P. Robertson, M. Douglas, B. Maruyama, "Political discourse on social networking sites: Sentiment, in-group/out-group orientation and rationality", *Information Polity*, Vol. 18, No. 2,pp. 107-126,2013.

[16] L.A.Hughes and L.Palen "Twitter adoption and use in mass convergence and emergency events", *International Journal of Emergency Management*, pp. 248-260, 2009.

[17] A. Jungherr,P. Jrgens and H. Schoen, "Why the Pirate Party Won the German Election of 2009", *Social Science Computer Review*, Vol.30,pp.229-234,2012.

[18] F.A.Nooralahzadeh, "2012 Presidential Elections on Twitter--An Analysis of How the US and French Election were reflected in Tweets", in *19th International Conference on Control Systems and Computer Science* [CSCS], Bucharest, pp.240-248,2013.

[19] B. O. Connor, R. Balasubramanyan, B. R.Routledge and N. A.Smith , "From tweets to polls: Linking text sentiment to public opinion time series" in: *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Washington, DC, May 2010.

[20] A. Tumasjan, T.O. Sprenger, P.Sandner, and I.Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment" in *International AAAI Conference on Weblogs and Social Media* 2010.

[21] D. J.Hopkins and G King, "A Method of Automated Nonparametric Content Analysis for Social Science" *American Journal of Political Science*, vol.54, pp.229–247,2010.

[22] N. A.Diakopoulos, D. A.Shamma , "Characterizing debate performance via aggregated twitter sentiment" in: *CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* ,pp. 1195-1198,2010.

[23] M.Choy , L. F. M.Cheong , M. N.Liak, K. P.Shung , "A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction" *Research Collection School Of Information Systems* ,2012.

[24] D.Gayo-Avello , "Don't turn social media into another'literarydigest'poll" in: *Communications of the ACM* vol.54, pp 121-128,2011.

[25] A.Bermingham and A. F. Smeaton , "On using Twitter to monitor political sentiment and predict election results"in *International Joint Conference for Natural Language Processing (IJCNLP)*, Chiang Mai, Thailand,pp.56-63,2011.

[26] P. T. A. Metaxas, "How (not) to predict elections: in *IEEE third international conference on social computing (SocialCom)* ,pp. 165-171,2011.

[27] M. D. Conover, B.Goncalves, J.Ratkiewicz, A.Flammini, and F.Menczer , "Predicting the political alignment of twitter users"in *IEEE third international conference on social computing (socialcom)* ,pp. 192-199,2011.

[28] A. D. Maynard and A Funk, "Automatic detection of political opinions in Tweets: in *ESWC 2011 Workshops*, Heraklion, Greece,pp 88-99,2011

[29] E. T. Sang, "Predicting the 2011 dutch senate election results with twitter"in *Workshop on Semantic Analysis in Social Media*, 2012,pp. 53-60.

[30] A. O. Larsson and M Hallvard"Studying political microblogging: Twitter users in the 2010 Swedish election campaign" *New Media & Society*, vol.14,729-747,2012.

[31] M. C. Choy , "US Presidential Election 2012 Prediction using Census Corrected Twitter Model" *Computers and Society*,2012.

[32] M Skoric , N Poor, P Achananuparp, E P Lim, , and J Jiang, "Tweets and votes: A study of the 2011 singapore general election"in *System Science (HICSS), 45th Hawaii International Conference on System Science IEEE*,pp. 2583-2591,2012.

[33] E.Tjong, K Sang, J Bos, Predicting the 2011 Dutch Senate Election Results with Twitter" in *EACL 2012Workshop on Semantic Analysis in Social Networks*, Avignon, France,2012.

[34] F .B. Marquez, D Gayo-Avello, M Mendoza and B Poblete "Opinion Dynamics of Elections in Twitter "in *LA-WEB '12 Proceedings of the 2012 Eighth Latin American Web Congress*,pp. 32-39,2012.

[35] J. M.Soler, F Cuartero, M Roblizo, "Twitter as a Tool for Predicting Elections Results" in *ASONAM*, IEEE/ACM International Conference, 2012, pp.1194 – 1200,2012.

[36] H. Wang, C. Dogan, A. Kazemzadeh ,F. Bar and S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle" in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Republic of Korea, pp. 115–120,2012.

[37] D. Contractor, T. A.Faruquie "Understanding Election Candidate Approval Ratings Using Social Media Data, in *WWW 2013 Companion*, Rio de Janeiro, Brazil,2013.

[38] M. P. Cameron, *Can Social Media Predict Election Results? Evidence from New Zealand*, Working Paper in Economics 13/08, University of Waikato, 2013.

[39] A. C. Ceron, "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France" *New Media & Society*, vol. 16, 340-358, 2013.

[40] M. S. Gaurav, "Leveraging candidate popularity on Twitter to predict election outcome" in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*.(ACM) 2013.

[41] F .T. Wong, "Media pundits and the US presidential election: Quantifying political leanings from tweets" in *Proceedings of the International Conference on Weblogs and Social Media*, 2013.

[42] E. Sanders , "Relating Political Party Mentions on Twitter with Polls and Election Results" in *Proceedings of the 13th Dutch-Belgian Workshop on Information Retrieval*,pp. 68-71,2013.

[43] C. B. Fink, "Twitter, Public Opinion, and the 2011 Nigerian Presidential Election" in *SocialCom*, International Conference on Social Computing IEEE,pp.311-320,2013.

[44] H. D. Kim, M. Castellanos , M. Hsu , C. X. Zhai, T. Rietz , D. Diermeier , "Mining Causal Topics in Text Data: Iterative Topic Modeling with Time Series Feedback", in *CIKM'13*, San Francisco, CA, USA,2013.

[45] A. Bakliwal, J. Foster, J. V.Puil, R. O'Brien, L .Tounsi and M. Hughes, "Sentiment Analysis of Political Tweets: Towards an Accurate Classifier" in: *LASM*, Workshop on Language in Social Media ,Atlanta, Georgia,pp. 49–58,2013.

[46] N. A.Thapen, M MGhanem, “Towards Passive Political Opinion Polling using Twitter” in *BCS SGAI Workshop on Social Media Analysis* , Cambridge, UK.2013.
[47] C.Vaccari, A Valeriani , P.Barberá, R.Bonneau “Social media and political communication: a survey of Twitter users during the 2013 Italian general election” *Rivistaitaliana di scienzapolitica*, vol.43,pp. 325-355,2013.
[48] M.Vergeer, L Hermans, “Campaigning on Twitter: Microblogging and Online Social Networking as Campaign Tools in the 2010 General Elections in the Netherlands” *Journal of Computer-Mediated Communication* ,vol.18 ,pp. 399–419,2013.
[49] F. M. Wong, “Quantifying Political Leaning from Tweets and Retweets” in *ICWSM*,2013.
[50] M. Song , C M Kim, Y K Jeong, “Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter”,*IEEE Intelligent Systems* ,vol.1,pp.1541-1672,2013.
[51] A. R.Makazhanov, “Predicting political preference of Twitter users” *Social Network Analysis and Mining*, vol.1, pp.1-15,2014.
[52] G.Ifrim, B. Shi, I.Brigadir , “Event Detection in Twitter using Aggressive Filtering and Hierarchical Tweet Clustering”,in *SNOW 2014 Data Challenge*, Seoul, Korea, 2014.
[53] P.Mehndiratta , S Sachdeva, P Sachdeva, Y Sehgal “Elections Again, Twitter May Help!!! A Large Scale Study for Predicting Election Results Using Twitter” in *Third International Conference, BDA 2014*, New Delhi, India,pp. 133-144,2014.
[54] P.Barbera and G.Rivero, “Understanding the Political Representativeness of Twitter Users”, *Social Science Computer Review*, vol.1, pp.56-67, 2014.
[55] C.Vaccari, A Valeriani, P Barberá, R Bonneau, J T Jost, J Nagler and J A Tucker , “Political Expression and Action on Social Media: Exploring the Relationship Between Lower- and Higher-Threshold Political Activities Among Twitter Users in Italy” *Journal of Computer-Mediated Communication*,vol.20,pp. 221–239.2015.
[56] Elvyna Tunggawan,Yustinus Eko Soelistio, “And the Winner is ...: Bayesian Twitter-based Prediction on 2016 U.S. Presidential Election” ,*International Conference on Computer, Control, Informatics and its Applications*.pp.33-37,2016.
[57] B. Narendra, Mr. K. Uday Sai, Mr. G. Rajesh, Mr. K. Hemanth, Mr. M. V. Chaitanya Teja, Mr. K. Deva Kumar “Sentiment Analysis on Movie Reviews: A Comparative Study of Machine Learning Algorithms and Open Source Technologies” *I.J. Intelligent Systems and Applications*, vol.8, pp.66-70,2016 .
[58] Pooja Wadhwa, M.P.S Bhatia, “Discovering Hidden Networks in On-line Social Networks”, I.*J. Intelligent Systems and Applications*, vol.5, pp.44-54, 2014.
[59] David M. Blei, Latent Dirichlet Allocation, Journal of *Machine Learning Research*, vol.3, pp.993-1022, 2003.
[60] Mayy M. Al-Tahrawi, “Arabic Text Categorization Using Logistic Regression” *I.J. Intelligent Systems and Applications*, vol.6,pp. 71-78,2015.
[61] Rasim M. Alguliyev,Ramiz M. Aliguliyev and Irada Y. Alakbarova, “Extraction of Hidden Social Networks from Wiki-Environment Involved in Information Conflict”, *I.J. Intelligent Systems and Applications*, vol. 2,pp. 20-27,2016.
[62] Y Bao, C Quan,L Wang,F. Ren , “The Role of Pre-processing in Twitter Sentiment Analysis”, in *ICIC*,*Taiyuan*, China,pp.615-624,2014.
[63] Indrajit Mukherjee,Jasni M Zain,P. K. Mahanti, “An Automated Real-Time System for Opinion Mining using a Hybrid Approach”,*I.J. Intelligent Systems and Applications*, vol.7,p. 55-64,2016.
[64] Ibrahim S. I. Abuhaiba,Hassan M. Dawoud, “Combining Different Approaches to Improve Arabic Text Documents Classification” *I.J. Intelligent Systems and Applications*, vol.4,pp. 39-52,2017.

Authors’ Profiles

Vinay K. Jain received his B.E. in Computer Science and Engineering in 2009 from Rajiv Gandhi Proudyogiki Vishwavidyala, Bhopal, India and received his M.Tech in Computer Science and Engineering from Jaypee University of Engineering and Technology, Guna, India in 2012. Now, he is pursuing his Ph.D. degree from Jaypee University of Engineering and Technology, Guna, M.P., India. He has published several papers in peer-reviewed International and Scientific Journals. He is also serving as reviewer for several Science Citation Indexed and Scopus Indexed International Journals.

Shishir Kumar in working as Professor the Department of Computer Science and Engineering at Jaypee University of Engineering and Technology, Guna, M.P., India. He has earned PhD in Computer Science in 2005. He has 14 years of teaching and research experience.

How to cite this paper: Vinay K. Jain, Shishir Kumar, "Towards Prediction of Election Outcomes Using Social Media", International Journal of Intelligent Systems and Applications(IJISA), Vol.9, No.12, pp.20-28, 2017. DOI: 10.5815/ijisa.2017.12.03