International Journal of Engineering & Technology, 7 (4.5) (2018) 366-369

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Election Prediction Using Big Data Analytics-A Survey

Abdul Manan koli[1]*, Muqeem Ahmed[2]

[1]Research Scholar (Manuu) A central University, Hyderabad
[2]Assistant Professor (Manuu) A central University, Hyderabad
Corresponding author E-mail:[1]*manankohli14@gmail.com, [2]muqeem.ahmed@gmail.com*

Abstract

Social media has received much attention due to it's real-time and interactive nature for political discourse, especially around election times. Recently studies have explored the power of social media platforms such as Twitter or Facebook, on recording current social trends and predicting the voting outcomes of an area. These social media generate a large amount of raw data that can be used in decision making for election predictions. This tremendously generated data is referred to as “Big data”. After scrutinized a lot of research work related to election prediction, a survey paper is presented in which every work related to election prediction using social media is incorporated. This paper is an attempt to review various tools, models, and algorithms used for the observation of campaign, discussion, prediction, and analysis of the election, and also suggest further tools and techniques for improvement.

Keywords: *Social Media, Twitter, Election Prediction, Hadoop, Big Data, Parameters.*

1. Introduction

The term “Big data” came into limelight when traditional database tools like “RDBMS” become incapable to handle large, unstructured data which is characterized by high volume, velocity, and variety [1]. Extracting needful information from this big data is one of the main challenges for both analysts and database. Every day approximately 2.5 quintillion bytes of data are created [2]. With this abundance of data, we can easily generate some meaningful information by applying suitable techniques to the data set [2]. Elections are held throughout the world covering almost all the entire nations. The election is a process by which general public can choose their representative by casting their votes. Every nation has different terms and norms for the election process. Social Media is an online application platform which facilitates interaction, collaboration, and sharing of content [44]. Both public, as well as political leaders, use social media like Twitter, Facebook, and Google+ etc for the campaign, discussion, prediction, and analysis of the election. These social media especially Twitter and Facebook generate an enormous amount of raw data which is very beneficial for both political parties and general public especially during election times. Hence political parties use social media because Politicians with higher social media engagement got relatively more votes within most political parties [45] and also it increases their campaign because fan base of leading political leaders increased with the onset of their digital campaign during the elections [46]. So Political parties, even in developing countries, make conscious efforts to manage social media properly during their campaigning phase [47]. Importance of social media comes into the limelight when Barrack Obama won 2008 Presidential election of America by using social media (Twitter) in his campaigning [48].

2. Related Work

Over the years, numerous works have been done related to election analysis and prediction system. But till date, we don’t find any suitable techniques and tools which can accurately predict the election outcomes. This paper aims at analyzing different data mining techniques and tools that have been introduced in recent years in the field of election analysis.

Wiji Arulampalam, et al. (2008). Framed a theoretical model to determine the allocation of funds between center and state of the Indian subcontinent. For their research work, they used log-linear model for grant and select 14 different states from 1974–75 to 1996–97. After examination they found that the states which are aligned to the centre in both the Vidhan Sabha and the Lok Sabha election (i.e. state and centre should have same ruling party) receive higher grants approximately (19.6%) then the state which are non-aligned, and those that are aligned with the centre in Vidhan Sabha but non-swing in the Lok Sabha is estimated to receive (13.9%) higher central grant then the states which are unaligned with both election types. [30]

Draper and Riesenfeld (2008) presented an interactive visualization technique that allows users to construct queries on Metadata sets, and view the results in short Spain of time. They purpose a novel interactive visualization for querying and analyzing tabular demographic data so that naive, as well as trained professional user, ascertains vast data in similar fashion and times. After proper scrutinizing it was revealed that this technique plays an effective role in analyzing opinion poll data. [31]

Campbell and Lewis-Beck (2008) investigated the possibility to predict the election outcome before the actual results. For this purpose, they investigate the previous research started from Lee sigleman (1979) to (2008) and came out with the conclusion that this method can be used in prediction but needs a lot of refinements in data collection methods, parameters used for prediction and also advancements in technologies. [32]

Claes H. De Vreese (2009) expanded the scope of second-rate political election campaigns of eight European countries for 2004 elections. After examine it was observed that during campaigns mostly political parties were very actively involved on burning issues of most European integration, and large no. of public meetings, more campaigning and actively using social media were organized.[33]

Tumasjan Andranik, et al. (2010) researchers traced the consequences of the data collected from twitter that can be used as an indicator in predicting the election outcome. To extract the sentiments from tweets, authors used linguistic Inquiry and word count tools and revealed that data obtained from twitter can be used in predicting the election results with possibly less time and minimum error rates. [9]

Brendan O'Connor, et al. (2010) measured sentiments of public opinion derived from polls conducted on the popular microblogging site Twitter. Data is collected from Twitter API which is used to demonstrate the consumer confidence and political opinion, and can also be used to analyze movements in the polls further. After classification and analyzing it was observed that this method predicts public opinion about polls with great accuracy rate, hence from its potential it is found that this technique can be used as supplements with traditional polls methods. [34]

Emre Toros. (2011) proposed a model for predicting elections in Turkey based on three theoretical premises. The equation for model is Change in vote share = "Economic conditions + Local election success + political structure" with Lewis-Beck's tools for quality score index and mean absolute error. After examine they found that a party namely AKP has a winning chance of 38.70% in contrast to other opposition parties who have less chance. [35]

Conover, et al. (2011) developed a method for predicting the political alignment of active Twitter user for 2010 U.S. midterm elections. They use support vector machine (SVM) trained on Metadata of Twitter user, with latent semantic analysis to identify the hidden information of vast data generated by the different user. For classification and communication network clustering algorithm was used. After mining, it was revealed that this method can be used for prediction of election analysis with (91%) accuracy rate.[36]

Harmanjit Singh, et al. (2012) used fuzzy-logic interface system for predicting the chance of winning or losing of an electoral candidate based upon nine different parameters. To determine the output Mamdani algorithm is used. After applying this method it was concluded that this technique plays a pivotal role in election prediction. [37]

Harmanjit Singh, et al. (2013) designed a simple Graphical user interface (GUI) model based upon fuzzy cognitive maps which is used for predicting the winning chances of a candidate based upon ten different variables. After proper analysis, it was admitted that the proposed model can be used for prediction of the result before the actual result announced, but more parameter needs to be added in order to enhance the better predictability. [38]

Dr. Mahmood Tariq, et al. (2013) used Twitter data in the prediction of election outcomes. Data was collected using a website called Twimemachine. CHAID decision tree, Naive Bayes and Support Vector Machine were used in prediction and the Rapid Minor tool was used in the mining of Metadata obtained from tweets respectively. After examining it was analyzed that CHAID algorithm has better results in predictions as compared to other algorithms. [12]

Hummel and Rothschild, (2013) developed a model which can be used to predict the election outcomes of Presidents, Governor and members of legislative-based upon the different parameter. By using linear regression the researcher concludes that this model can be best suited for prediction of an above-stated politician with winning the chance of (90% for president), (82.5% for senatorial) and (79.1% for Governor) with fewer error rates respectively. So this model can be used for election outcome but the main disadvantages of this model is that the parameters used in it are mainly from economic and political sphere other parameter needs to be added in such analysis. [39]

Wani and Alone, (2014) analyzed the impact of social media on heterogeneous Indian politically system. In India, 65% populations are below 35 years of age and use social media as a platform in sharing their views, ideas, communication, sentiments etc and these factors play a key role in analyzing election process. The researcher uses Twitter data in surveying the user opinion about political parties' and extraction of actual relevant data is done using topic modeling techniques with KNN algorithm. [40]

Min Song, et al. (2014) used various data mining techniques on Twitter data and predict 2012 Korea Presidential election. They use multinomial topic modeling, network analysis, and co-occurrence retrieval techniques for extraction of necessary information from vast twitter dataset. After analysis, it was revealed that this technique can be used to mine dynamic social trends and content-based networks generated in Twitter. [41]

Conway A. Bethany, et al. (2015) researchers included the impact of another social network apart from tweeter and facebook. By collecting tweets and article published in top newspapers using QDA miner and word state techniques, it was revealed that there exists a nonrandom relationship between both of them, however, Twittersphere is an emerging trend in the election campaign but traditional media also plays a vital role. [4].

Singhal Kartik, et al. (2015) Proposed a novel approach based on a semantic and context-aware rule to detect public opinion for prediction of 2014 Delhi (India) elections. A hybrid approach to Lexicon based and rule-based sentiments analysis was used. For counting the sentiment score of the word and for removing extra word which is not related to sentiments Stanford Parser tool was used [11].

Khatua Aparup, et al. (2015) discussed the Idea of Twitter data in predicting election results of large and politically diversified country India. The main challenge was collection of the vast amount of hybrid data because of both national, as well as regional parties, participated in the general election. Lexicon method for sentimental analysis and OLS Regression model for vote swing with Mean absolute error for error detection was used [15]

Tsakalidis Adam, et al. (2015) Gave significance of Twitter data in predicting the election results of three nations. The challenge includes gaining access to data, capabilities needed to work with a large dataset of three nations and finally predict the better election results as compared to others models. For predictions and analysis linear regression, Gaussian Process and sequential minimal optimization for regression were implemented using WEKA tool with mean absolute error and mean squad error. [20]

Kangan Vadim, et al. (2015) Researchers used Twitter data in predicting the election results using diffusion estimation model and sentiment analysis algorithm. Further, they proposed the Indian Election Tweet database (IET-Dba proposed model) to forecast the better elections outcome as compared to other methods like (Pew Research Poll). And using such information valuable intelligence can be obtained. [21]

Jagdev Gagandeep and Kour Gagandeep (2016) used big data technologies like Hadoop and Map Reduce for the analysis of Metadata. Hadoop tool was used for acquiring and storage of vast heterogeneous data and Map reduce algorithm for sorting and processing, and relevant information of election was extracted in much shorter time with less cost consumption as compared to traditional database tool like RDBMS [17].

Zheng Xie, et al. (2016) collected online and offline data of public opinions in predicting of Taiwan Presidential elections 2016. Researchers applied Kalman Filter techniques for signal processing and moving average model for real-time burst detection on vast data in predicting Taiwan President election with an error rate less than 3%. [22].

Daniel, et al. (2016) analyzed the election results by comparing sentiments analysis of Twitter data with traditional opinion polls during Brazilian President Election 2014. Discover Text and Rapid miner tool were used with natural language processing techniques to extract relevant information and revealed that two were very similar in prediction with less accuracy of sentiment analysis contrast to traditional polls. [6]

Sharma Parul and Moh Sheng Teng (2016) gave the significance of using sentiment analysis in predicting election outcomes using Twitter data collected in Hindi language formats (tweets). They used Dictionary based, Naive Bayes, and support vector machine algorithms with H-SWN (Hindi-sentwordnet) in their work. After mining the dataset of Hindi tweets they reveal it can be used in predicting election results. [8]

Yu Wang et al. (2016) attempted to characterize the users who have left the presidential candidates during the election campaign on Twitter. They took four parameters for their research work (i.e. social capital, gender, age, and race), and use convolutional neural network and Face++ API tool for classification and analysis. After mining the data set of twitter they concluded that female user is more prone to unfollower than the male user. [42]

Goyal Shubham (2017) extracted writer comments or reviews from Twitter using hybrid approach by using KNN algorithm and Naive Bayes algorithm with Twitter API in their research work. After extraction and mining, they reveal that social network site like Twitter can be used in predicting the sentiments of peoples. [7]

Suarez Hernandez A, et al. (2017) presented a mood analysis methodology for predicting social sentiment of political events using Twitter data. By pre-classifying tweets with positive and negative labels with Naive Bayes classifier, it was revealed that proposed method is useful in observing online user behavior towards political issues during elections. [19]

Gourav, et al. (2017) used 'tweets' and analyzed the Prime Minister candidates from Indian political diplomats between Mr. Modi and Mr. Kejriwal. They applied data mining techniques and text mining methods to study the tweets. Analyses of these tweets were conducted in R studio using Natural Language Processing methods and Twitter search API for predictions of elections results [5].

Safiullah, et al, (2017) investigated the predicting power of social media especially twitter in election outcomes. They use regression analysis techniques for data analysis with root mean squared error. After examining dataset properly it was revealed that social media buzz especially twitter can be used as a healthy indicator in election prediction. [43]

Singh Prabhsimran and Singh Ravinder, (2017) the main aim of this paper is to find either the data obtained from the social networking site i.e Twitter can be used for election prediction. And what are the major hurdle faced by this predictions. For this analysis purpose they took previous research work related to election prediction using Twitter data of different nations. After proper mining they revealed that countries in which the internet user are above (80%) suited fit for this analysis. On the other hands countries which have internet user less than (80%) were not suitable for this predictions. So it can be quoted that Twitter can be used as indicator but not strong parameter for predictions. [52]

3. Conclusion

Literature survey showed that there is no single research work that can precisely anticipate the decision results of elections. Although many of them had made an effective attempt in election result predictions but they are post-hoc analysis. Most basic downside which is gliding practically in each exploration work is neglecting the democracy. Because social media like Twitter and Facebook include selected portion of the population. The common drawback which is floating almost in every research work is ignoring the majorities which don't utilize the social locales like Facebook and Twitter. In developing nations twitter and facebook is mostly used in urban areas and such countries have less number of populations residing in urban sites as compared to rural areas. So social media especially twitter and facebook may produce biased result. It has also been analyzed that there is no single classifier and tool which give consistent and accurate results for all types of elections and hence predict the election results with possibly less time and minimum error rates. The current machine learning algorithms being used in the prediction systems are able to take simpler decisions but lack in multi-level prediction, these algorithms are becomes slower when they have large datasets to be trained and tested upon. Further investigation is still needed because the availability of huge amounts of election data leads to the need for powerful data analysis tools to extract useful knowledge. From this study, it can be concluded that parameters and methodology can be further increased so that correct accuracy can be achieved.

4. Future work

To enhance the consequences of social sites some efficient algorithms with hybrid mechanisms should be incorporated. Advanced tools and techniques like Hadoop, Hive, spark and Rapid miner play vital role in the prediction of election results hence should be added too. Platform like, Python will be used to develop the system which will lessen the development time and also it contains several libraries which supports big data and advanced machine learning algorithms. Emphasis should be given to cover an entire geographic area of the particular nation instead of few states. Sentimental analysis, image base and text-based analysis methods should be used so that whole public opinion polls may be added in prediction. The parameters used for election analysis and prediction should not be only from the economic and political spheres but whole environment should be kept in mind.

References

[1] Chacón D, (2013). "Why Big Data Is Important to You and Your Organization".

[2] Nasser T and Tariq RS "Big Data Challenges" (2015). Journal of Computer Engineering and Information Technology.

[3] Salla-Maaria Laaksonen, et al., Working the fields of big data: Using big-data-augmented online ethnography to study candidate-candidate interaction at election time. Journal of information technology & politics, 2017.

[4] Bethany Anee Conway et al.,The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. Article in Journal of Computer-Mediated Communication · May 2015.

[5] Dubey Gaurav, et al., Social media opinion analysis for Indian political diplomats. IEEE 2017.

[6] Daniel José Silva Oliveira, et al., Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls, Journal of Information Technology & Politics, 2016.

[7] Shubham Goyal. , Review Paper on Sentiment Analysis of Twitter Data Using Text Mining and Hybrid Classification Approach. International Journal of Engineering Development and Research (www.ijedr.org) 2017.

[8] Parul Sharma and Teng-Sheng Moh., Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter. , IEEE 2016.

[9] Andranik tumasjan, et al., Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. Association for the advancement of artificial intelligence, 2010.

[10] Zhihan Lv, et al.Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. IEEE 2016.

[11] Kartik Singhalj et al., Modeling Indian general elections: Sentiment Analysis of Political twitter data.

[12] Dr. Tariq Mahmood, et al., Mining Twitter Big Data to Predict 2013 Pakistan Election Winner. IEEE, 2013.

[13] Tapio Vepsalainen. et al., Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections. Elsevier 2017.

[14] dr gagan deep jagdev., et al., Excavating Big Data associated to Indian Elections Scenario via Apache Hadoop. International Journal of Advanced Research in Computer Science www.ijarcs.info , 2016.

[15] Khatua et al., Can #Twitter_Trends Predict Election Results? Evidence from 2014 Indian general election.IEEE 2015.

[16] Miss .Payal Rajkumar Rathi et al.,Big Data Analytics for Social Network--the Base Study, http://www.ijettjournal.org.2016.

[17] Gagandeep Jagdev and Amandeep Kaur; Analyzing and Scripting Indian Election strategies using Big Data via Apache Hadoop framework, IEEE 2016.

[18] Andy Januar Wicaksono et al. , A Proposed Method for Predicting US Presidential Election by Analyzing Sentiment in Social Media, 2nd International Conference on Science in Information Technology , IEEE 2016.

[19] Suarez Hernandez A, et al., predicting political mood tendencies based on twitter data.

[20] Tsakalidis Adam, et al. , predicting elections for multiple countries using twitter and polls. IEEE 2015.

[21] Kangan vadim, et al., using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election. IEEE 2015.

[22] Zheng xie, et al., Wisdom of fusion: Prediction of 2016 Taiwan Election with Heterogeneous Big Data. IEEE 2016.

[23] E. Sang and J. Bos, “Predicting the 2011 Dutch Senate Election Results with Twitter,” Proc. Workshop Semantic Analysis in Social Media, 2012, pp. 53–60.

[24] B. O’Connor et al., “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series,” Proc. 4th Int’l AAAI Conf. Weblogs and Social Media

[25] Twitter Statisitics, http://www.statisticbrain.com/twitter-statistics/, Last Updated: 5.7.2013

[26] C. Spengler, W. Werth, and R. Sigrist, “360o Touchpoint Management -How important is Twitter for our brand”, Marketing Review, St. Gallen, 2010.

[27] M. Ghiassi, J. Skinner, and D. Zimbra, “Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network”, Expert Systems with Applications, Vol. 40, Issue 16,pp 6266-6282, November 2013.

[28] Y. Bae and H. Lee, “Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers”, Journal of the American Society for Information Science and Technology, Vol. 63,Issue 12, pp. 2521-2535, December 2012.

[29] M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Kiem, L.E.. Haug,M. C. Hsu, “Visual Sentiment Analysis on Twitter Data Streams”, IEEE Symposium on Visual Analytics Science and Technology, October 23-29, Providence, Rhode Island, USA.

[30] Wiji Arulampalam, et al. “Electoral goals and center-state transfers: A theoretical model and empirical evidence from India. (2008). Elsevier 2008

[31] Geoffrey M. Draper and Richard F. Riesenfeld “Who Votes for What? A Visual Query Language for Opinion Data. (2008). IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 14, NO. 6, NOVEMBER/DECEMBER 2008

[32] James E. Campbell and Michael S. Lewis-Beck US presidential election forecasting: An introduction. (2008). Elsevier 2008

[33] CLAES H. DE VREESE, “Second-Rate Election Campaigning? An Analysis of Campaign Styles in European Parliamentary Elections” (2009). Journal of Political Marketing, 8:7–19, 2009

[34] Brendan O’Connor, et al. “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series” (2010). Association for the Advancement of Artificial Intelligence 2010.

[35] Emre Toros. “Forecasting elections in Turkey” (2011). International Institute of Forecasters. Published by Elsevier 2011.

[36] Michael D. Conover, et al. “Predicting the Political Alignment of Twitter Users” (2011). IEEE International Conference on Social Computing 2011.

[37] Harmanjit Singh, et al. “Election Results Prediction System based on Fuzzy Logic” (2012). International Journal of Computer Applications (0975 – 8887) Volume 53– No.9, September (2012)

[38] Harmanjit Singh, et al. “Fuzzy Cognitive Maps Based Election Results Prediction System” (2013). https://www.researchgate.net/publication/274365593 2013

[39] Patrick Hummel and David Rothschild, “Fundamental Models for Forecasting Elections” (2013). researchdmr.com 2013.

[40] Gayatri Wani and Nilesh Alone, “A Survey on Impact of Social Media on Election System” (2014). International Journal of Computer Science and Information Technologies, Vol. 5 (6), 2014.

[41] Min Song, et al. “Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter”, (2014). IEEE Intelligent Systems, 2014

[42] Yu Wang et al. “Voting with Feet: Who are Leaving Hillary Clinton and Donald Trump”. (2016). IEEE International Symposium on Multimedia 2016.

[43] Md Safiullah, et al, “Social media as an upcoming tool for political marketing effectiveness” (2017). Elsevier 2017.

[44] (Palmer and Koening-Lewis “An experiential, social network-based approach to direct marketing” (2009),

[45] (Effing et al., “Social media and political participation: are Facebook, Twitter and Youtube democratizing our political system?” (2011)

[46] (Singh, 2014).

[47] (Ahmed, Saifuddin et al. “My name is Khan: the use of Twitter in campaign for 2013 Pakistan General Election” 2014).

[48] (Tumasjan et al. “Predicting elections with twitter: What 140 characters reveal about political sentiment”2010).

[49] From database to Big data, By : Ssm Madden – MIT http:// ieeexplore.ieee.org/stamp/stamp.jsp? tp=& arnumber= 6188576

[50] Shira Fano and Debora Slanzi , Using Twitter Data to Monitor Political Campaigns and Predict Election Results. Springer international publishing AG 2017

[51] Muhammad Zubair Asghar et al., T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme, wileyonlinelibrary.com/journal/exsy

[52] Prabhsiran SinghRavinder Singh Sawhney, “Influence of Twitter on Prediction of Election Results”. Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, vol 564. Springer, Singapore