JOURNAL OF AERONAUTICAL MATERIALS
ISSN: 1005-5053

Vol. 43, Issue-01, 2023
pp. 290-298

AN ANALYSIS OF ELECTION PREDICTION USING SOCIAL MEDIA DATA NETWORK: A REVIEW

Sadique Nayeem
Assistant Professor, Department of CSE, Sitamarhi Institute of Technology, Sitamarhi

Dr. Savya Sachi
Assistant Professor, Department of Information Technology, L. N. Mishra College of Business Management, Muzaffarpur

Ramesh Kumar
Assistant Professor , Department of Computer Science & Engineering,
B. P. Mandal College of Engineering, Madhepura

Abstract- The emergence and widespread use of modern social media (SM), including social networks like Facebook, Twitter, and Instagram, has changed how lawmakers interact with the public and conduct political campaigns. (SNs). Due to the intrinsic strengths of SM, including the vast amount of data that can be accessed in real time, a new area of study has emerged that focuses on using SM data to forecast election results. Despite the fact that numerous studies have been done in the last ten years, the findings are frequently disputed. In terms of methodology, we conducted a systematic literature review, analysing the quantity and quality of publications, the electoral context of studies, the main approaches to and characteristics of successful studies, as well as their main strengths and challenges, and comparing our findings to previous reviews. In this context, the purpose of this article is to investigate and summarise how research on predicting elections based on SM data has developed since its inception, to describe the state of the art and practice, and to identify research opportunities within this field.
Keywords— Emergence, Widespread Modern Social Media (SM), Election Prediction, Social Media Data Network.

INTRODUCTION

Social media can be defined as a collection of internet-based applications that expand the ideological and technological foundations of Web 2.0 and that permit the creation and exchange of user-generated content. Social networking sites are the interface between people and social media, and for many the "Internet" is synonymous with social networking sites. One of the most interesting characteristic of social media is represented by the term "user-generated content", which refers to different forms of media content, publicly available and created by end users. Today youth is increasingly using numbers of social sites such as Facebook, Twitter, WhatsApp, Blog and LinkedIn as platforms for communicating with friends, family and work colleagues. Social Media‟s quick development shows its influence on society and is a crucial part of the advancement of information and communication technologies. Given its popularity, social media could be used as an influential channel for opinion leading, including agenda-setting and public opinion formation.

Election campaigns fundamentally rely on communication. Over the last decade, changes in the communication environment due to innovations in digital technologies, which themselves accompanied a process of modernization and professionalization of electoral competition have forced political elites to adopt and integrate in their campaigns increasingly sophisticated digital communication practices. Political parties and candidates embraced new online tools as part of their campaign communication Social networking sites like Facebook, microblogs like Twitter and video-sharing sites like YouTube have not only given politicians a powerful avenue for interacting with a more demanding citizenry, but also have allowed them to offer more personalized images to the public and have given less resourceful parties the opportunity to match well-funded campaigns in sophistication, using creative and relatively inexpensive strategies. Candidates, members of parliament, and local committee members worldwide are now providing information about their policy positions, invite followers to campaign events or meetings on Facebook, and interact with their constituencies "on the go" and through short messages on Twitter rather than long and time-consuming posts on their blogs or websites. In India, there was a significant change in the General Elections 2014 from the General Elections 2009; this was the change in the role played by the social media during the elections. It has been observed worldwide that the democracies have been engaging in dialogues with the public over the social media. An election is a most important part in the democracy. it's the most instrument of democracy wherever the voters communicate with the representatives. Due to their important role in politics, there always has been a big interest in predicting an election outcome. It is the main instrument of democracy where the citizens communicate with the representatives. One vital component in an election is that the election polls/survey.

Lately, it is observed that traditional polls may fail to make an accurate prediction. The scientific community has turned its interest in analyzing web data, such as blog posts or social networks' users' activity as an alternative way to predict election outcomes, hopefully more accurate. Furthermore, traditional polls are too costly, while online information is easy to obtain and freely available. This is an interesting research area that combines politics and social media which both concern today's society. It is interesting to employ technology to solve modern-day challenges. Social media has become the most popular communication tool on the internet. Hundreds of millions of messages are being posted every day in the popular social media sites such as Twitter and Facebook. Stated in their paper that social media websites become valuable sources for opinion mining because people post everything, from the details of their daily life, such as the products and services they use, to opinions about current issues such as their political and religious views. The social media providers enable the users to express their feelings or opinions as much as possible to increase the interaction between the users and their sites. This means that the trend on the internet is shifting from the quality and lengthy blog posts to much more numerous short posts that are posted by a lot of people. This trait is very valuable as now we can collect different kind of people's opinions or sentiments from the social web. The use and influence of social media in politics have been recognized by many researchers and political parties in making electoral predictions and devising future campaign strategies. In literature mostly work has been done on twitter and with sentiment analysis. But it is important to include all social media indicators into the account and also use an efficient method for sentiment analysis. It is also required to develop an optimized solution for sentiment analysis of user generated contents. Lastly combining data extraction, social

media indicators & sentiment analysis on various social media platform is also a critical task. Proposed research work will develop a framework to provide solution to all these problems.

LITERATURE REVIEW

According to Kaplan and Haenlein "Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content".

Types	Examples
Social Networks	Facebook, LinkedIn, MySpace, Googleplus
Blogs	Blogger, WordPress
Microblogs	Twitter, Tumblr
Social News	Digg, Reddit
Social Bookmarking	Delicious, StumbleUpon
Media Sharing	Instagram, Youtube
Wikies	Wikipedia
Review Sites	Yelp, Tripadvisor

Table 1: Type of social media platforms

Social Media consists of the following elements: openness, sharing, networking, communication, togetherness, co-creation or user-generated content. These characteristics are important for determining whether or not the future can be predicted with Social Media data. The most popular Social Networking Sites in the United States in March 2015 (based on market share of visits) were Facebook as number one, followed by Youtube, Google Plus and then Twitter. Table below represents various social media platforms:

Social Media Key Performance Indicators (KPIs)- Following are social media indicators used for analysis of social media platforms:

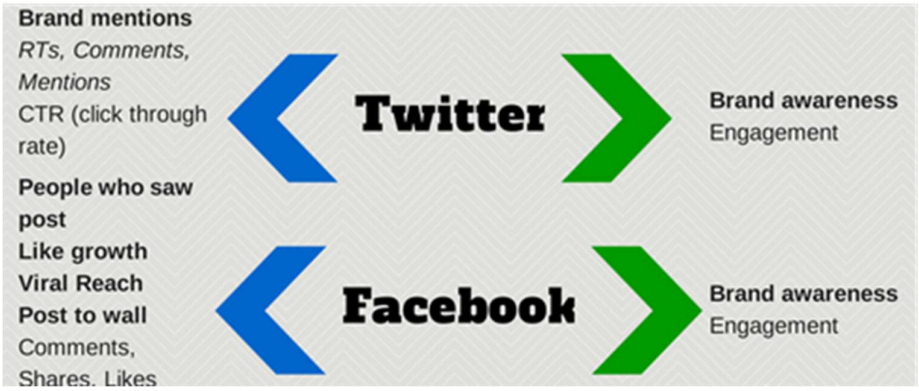

Fig 1: Social Media KPIs

Technique used to predict electoral result are classified into -volumetric analysis, sentiment analysis and social network analysis.

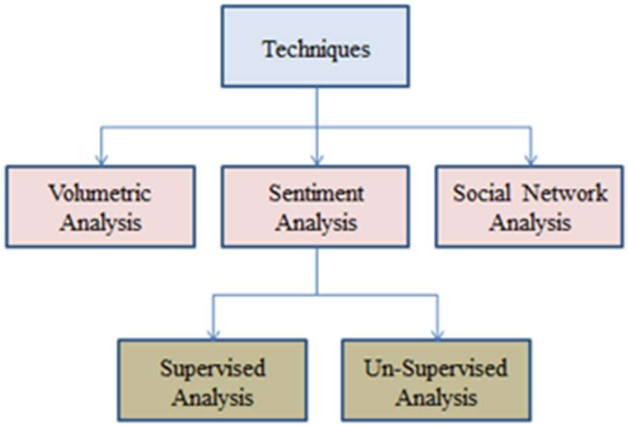

Fig 2: Technique used to predict electoral result

Sentiment analysis is the process of detecting the contextual polarity of the text. It determines whether given text is positive, negative or neutral. It is otherwise called as opinion mining too, since it derives the opinion or attitude of the speaker. In general, sentiment analysis has been investigated mainly at three levels. In document level the main task is to classify whether a whole opinion document expresses a positive or negative sentiment. This level of analysis assumes that each document expresses opinions on a single entity. In sentence level the main task is to check whether each sentence expressed a positive, negative, or neutral opinion. This level of analysis is closely related to subjectivity classification, which distinguishes objective sentences that express factual information from subjective sentences that express subjective views and opinion. Document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself.

Vepsäläinen et al has analyze the degree to which Facebook Likes could be used to predict the outcome of the 2015 Finnish parliamentary election. They found that the prediction based on Facebook Likes was less accurate than using incumbency and traditional polling on whether a candidate would be elected.

Prabhsimran et al has consider the Twitter data related to the 2017 Punjab (a state of India) assembly elections and applies different social media analytic techniques on collected tweets to extract and unearth hidden but useful information. In addition to this, we have employed machine learning algorithm to perform polarity analysis and have proposed a new seat forecasting method to accurately predict the number of seats that a political party is likely to win in the elections.

Marco et al revealing the political orientation of a Twitter user in the context of the 2016 Italian Constitutional Referendum. After having collected and processed over 1,200,000 tweets, they had classified them as YES-oriented, NO-oriented or UNCERTAIN, by exploiting the Naive Bayes Multinomial Text classification algorithm.

Parul et al used Twitter Archiver tool to get tweets in Hindi language. They performed data (text) mining on 42,235 tweets collected over a period of a month that referenced five national political parties in India, during the campaigning period for general state elections in 2016. They had made use of both supervised and unsupervised approaches. They utilized Dictionary Based, Naive Bayes and SVM algorithm to build our classifier and classified the test data as positive, negative and neutral. The results of the analysis for Naïve Bayes was the BJP (Bhartiya Janta Party), for SVM it was the BJP (Bhartiya Janta Party) and for the Dictionary Approach it was the Indian Nathional Congress. SVM predicted a 78.4% chance that the BJP would win more elections in the general election due to the positive sentiment they received in tweets. As it turned out, BJP won 60 out of 126 constituencies in the 2016 general election, far more than any other political party as the next party (the Indian National Congress) only won 26 out of 126 constituencies.

Loris et al presents a methodology, called IOM-NN (Iterative Opinion Mining using Neural Networks), for discovering the polarization of social media users during election campaigns characterized by the competition of political factions. The methodology uses an automatic incremental procedure based on feed-forward neural networks for analyzing the posts published by social media users. Starting from a limited set of classification rules, created from a small subset of hashtags that are notoriously in favor of specific factions, the methodology iteratively generates new classification rules. Such rules are then used to determine the polarization of people towards a faction.

Khatua et al investigate the complex nature of these mix tweets in a multi-party context, and argue mix tweeting patterns of users implicitly capture their political opinions. They predict the political leaning of users based on their mix tweeting patterns in the context of the 2014 Indian General Election. They had agglomerated 2.4 million tweets from 0.15 million unique users. Next, They had employ a multinomial logit regression model to test the hypothesized causal relation between mix tweeting patterns and the political leaning of users. Additionally, They also employ neural network-based algorithms to predict political

Lei Wang et al presented a new method for election prediction based on Twitter data analysis is proposed and applied to predict the 2017 French Election. Researcher states that in Twitter based election prediction it is critical to extract informative keywords or features reflecting true sentiment of voters. In addition, traditional prediction models may not be suitable for the data from social networks.

Onur Varol et al determines that early detection, occurring immediately at trending time, is a more challenging problem due to the minimal volume of activity data that is available prior to trending. A supervised learning framework exploits hundreds of time-varying features to capture changing network and diffusion patterns, content and sentiment information, timing signals, and user meta-data. We explore different methods for encoding feature time series. Using millions of tweets containing trending hashtags, we achieve 75% AUC score for early detection, increasing to above 95% after trending. They uses KNN classifier. They evaluate the robustness of the algorithms by introducing random temporal shifts on the trend time series. Feature selection analysis reveals that content cues provide consistently useful signals; user features are more informative for early detection, while network and timing features are more helpful once more data is available. This work represents an important step toward the

automatic detection of campaigns. The problem is of paramount importance, since social media shape the opinions of millions of users in everyday life. Further work is needed to study whether different classes of campaigns (say, legitimate advertising vs. terrorist propaganda) may exhibit characteristics captured by distinct features.

O'Connor et al. in presented the feasibility of using Twitter data as a substitute and supplement for traditional polls. Subjectivity lexicon is used to determine opinion scores (i.e., positive and negative scores) for each message in the dataset. Then, the authors computed a sentiment score. Consumer confidence

and political opinion are analysed and found to be correlated with sentiment word frequencies in Twitter data. However, they do not describe any prediction method.

Tumasjan et al. in examined whether Twitter can be seen as a valid real-time indicator of political sentiment. The authors also found that the mere number of messages reflects the election result and comes close to traditional election polls.

Sang et al. in analysed Twitter data regarding the 2011 Dutch Senate elections. The authors presented that improving the quality of the document collection and performing sentiment analysis can improve performance of the prediction. However, the authors need to manually annotate political mes- sages to compute sentiment weight and only the first message of every user is taken into account. In addition, the method relies on polling data to correct for demographic bias.

Makazhanov et al. in proposed political preference prediction models based on a variety of contextual and behavioural features. The authors extract all interactions of the candidates, group them on a per-party basis, and build a feature vector for each group. Both a decision tree-based J48 and Logistic regression classifiers are utilized for each party. However, this method needs labelling of training examples for each user. The labelling of training set based on a set of users whose political preferences are known based on the ex-

plicit statements (e.g., "I voted XXX today!") made on the Election Day or soon after. Moreover, it does not predict the election outcomes. However, there are several works presented the problems on election prediction using Twitter data.

Jungherr et al. in presented that a lack of well-grounded rules for data collection and the choice of parties and the correct period in particular can cause the problems.

Metaxas et al. in concluded that Twitter data is only slightly better than chance when predicting elections. However, the authors described three necessary standards for predicting elections using

Twitter data: (1) it should be a clearly defined algorithm, (2) it should take into account the demographic differences between Twitter and the actual population, and (3) black-box methods should be avoided.

Gayo-Avello has criticized several flaws in. For example, there is not a commonly accepted way of counting votes in Twitter. Sentiment analysis is applied as a black-box and demographics are neglected. Nevertheless, the author has outlined some of the research lines for future works in this topic. For example, researches need to clearly define which are a vote and the ground truth; sentiment analysis is a core task and researches should acknowledge demographic bias.

BACKGROUND RESEARCH

(i) Elections Prediction is Increasing with SM Data - Modern social media (SM) platforms are new; Facebook, Twitter, and Instagram all debuted to the public in 2006 and 2010, respectively. SM started to be used in contemporary political activities and to be regarded as a source for election prediction just a few days after their debut. Tilton may be credited with one of the earliest initiatives targeted at forecasting election outcomes using data from SM. With the study question "Could Facebook be used to estimate the results of a student election?" he attempted to predict election outcomes of a connected society, in this case a university, in 2008, just two years after Facebook's public debut. According to the results, his model was successful in 21 out of 27 attempts to forecast where the candidates would finish in a particular election. Although Tilton's study is rarely referenced by other studies in the field possibly because it has nothing to do with formal political scenarios we think it is a very insightful preliminary investigation in this area. Nearly all subsequent studies reference two studies that can be regarded as seminal in the field of using SM data to predict political elections. (2010) Tumasjan et al. offered research on the German federal election of 2009. They gathered all tweets mentioning any of the six parties with representatives in the German parliament, or well-known party members, and compared the number of tweets to the election outcomes. According to their findings, "the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls," they wrote. The same year, O'Connor et al. discovered that "a relatively simple sentiment detector based on Twitter data replicates consumer confidence and presidential job approval polls" using a method that was enhanced by tweet sentiment detection. Based on these two studies, the majority of subsequent research conducted worldwide, including in the Netherlands, Italy, France, India, Indonesia, Colombia, Chile, and the United States, has primarily used the volume of tweets coupled with automatic sentiment detection.

(ii) Analysing earlier reviews- Some researchers attempted to compile the available information in this field due to the variety of approaches, which led to varied outcomes even in replications of the same approach in the same context. A systematic review was published in 2013 by Kalampokis et al. with the goal of understanding the predictive potential of SM outside of the electoral context. They discovered that the three major approaches were based on volume, sentiment, and user pro-filing after examining 52 studies, 11 of which dealt with election predictions. Additionally, the use of linear regression in predictive analysis was noted, but not in research pertaining to the political context. Additionally, they confirmed that 65% of lexicon-based approaches failed, with the percentage rising to 40% for studies that used sentiment-related variables to test SM predictive power. The absence of predictive analytics evaluation and the dubious outcomes of electoral prediction studies were highlighted as their final points. The same year, Gayo-Avello presented a study that we believe to be the initial analysis specifically on using social media to predict elections, with a concentration on Twitter. He came to the conclusion that "the presumed predictive power regarding electoral prediction has been somewhat exaggerated" after reviewing ten prior studies from 2010 to 2013. Furthermore, he stressed the importance of using more modern techniques for sentiment analysis and identified volume and sentiment analysis as the two primary approaches. He also added more difficulties to the list, including the reliance on arbitrary decisions made by researchers regarding keywords, parties, candidates, and the choice of the data collection

period, as well as issues with Twitter, including demographic and self-selection bias, as well as bias related to spam, deceptive propaganda, and astroturfing. Regression models could be a potential future direction, he said as he concluded the research.

CONCLUSION

During the last decade, the use social media has provided a virtual community where the users express their intention, opinion and communicate with others. Organizations and researchers are very much interested to investigate the polarity of the user's opinions for making predictions and planning of future plans. The use and influence of social media in politics have been recognized by many researchers and political parties in making electoral predictions and devising future campaign strategies. In literature mostly work has been done on twitter and with sentiment analysis. But it is important to include all social media indicators into the account and also use an efficient method for sentiment analysis. This research can also be used for other decision making activity like recommendation systems, pattern analysis, brand popularity etc.

REFERENCES

[1] "Social Media KPIs", https://w2smarketing.com/2014/07/07/measuring-your-social-media-and-a-kpi-cheat-sheet/

[2] A. Khatua and E. Cambria, Predicting political sentiments of voters from twitter in multi-party contexts, Applied Soft Computing Journal, 2020

[3] Chadwick, A., "The Hybrid Media System: Politics and Power". Oxford University Press, New York. 2014

[4] Gibson de Zúñiga, H., Molyneux, L. & Zheng, P. "Social Media, Political Expression, and Political Participation: Panel Analysis of Lagged and Concurrent Relationships" Journal of Communication 64, 2014

[5] Hinton, S. and Hjorth, L. "Understanding Social Media" 1st edition, SAGE Publications, UK, London, 2013

[6] Hong, S., and Nadler, D. Does the early bird move the polls?: the use of the social media tool'twitter'by us politicians and its impact on public opinion. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times , ACM, 2011, pp. 182–186.

[7] Kaplan, A.M and Haenlein, M. "Users of the world, unite! The challenges and opportunities of Social Media." Business Horizons, 2010, 53, 59-68

[8] Kaplan, A.M., Haenlein, M. "Users of the world, unite! The challenges and oppurtunities of Social Media". Business Horizons, 53, 59-68. 2010

[9] Lei Wang and John Q Gan, "Prediction of the 2017 French Election Based on Twitter Data Analysis" , IEEE 2017

[10] Loris belcastro, riccardo cantini, fabrizio marozzo Domenico talia and paolo trunfio, "Learning Political Polarization on Social Media Using Neural Networks", IEEE, 2020

[11] Marco Campanale, Enrico G. Caldarola, "Revealing Political Sentiment with Twitter: The Case Study of the 2016 Italian Constitutional Referendum", International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2018

[12] Muhammad Bilal, Abdullah Gani, Mohsen Marjani, Nadia Malik, "predicting elections: social media data and techniques", IEEE, 2019
[13] Neha S. Joshi, Suhasini A. Itkat, " A Survey on Feature Level Sentiment Analysis" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014,
[14] Onur Varol, Emilio Ferrara, Filippo Menczer, Alessandro Flammini, "Early Detection of Promoted Campaigns on Social Media", arXiv:1703.07518v1 [cs.SI] 22 Mar 2017
[15] Parul Sharma, Teng-Sheng Moh, "Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter", IEEE, 2016
[16] Pedro Saleiro, Sılvio Amir, Mario Silva, Carlos Soares, "POPmine: Tracking Political Opinion on the Web" , IEEE, 2015
[17] Prabhsimran Singh, Yogesh K. Dwivedi, Karanjeet Singh Kahlon, Annie Pathania, Ravinder Singh Sawhney, "Can twitter analytics predict election outcome? An insight from 2017 Punjab assembly elections", Elsevier, 2019, https://doi.org/10.1016/j.giq.2019.101444
[18] Statista , "U.S. Social Media visit share. Retrieved from Statista website: http://www.statista.com/statistics/265773/market-share-ofthe-most-popular-social-media-websites-in-the-us/", 2015
[19] Tapio Vepsäläinen, Hongxiu Li, Reima Suomi, "Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections", Government Information Quarterly ,Elsevier, 2019, http://dx.doi.org/10.1016/j.giq.2017.05.004
[20] Abel, F., Hauff, C., Houben, G.J., Stronkman, R., Tao, K.: Semantics + filtering + search = twitcident. exploring information in social web streams. In: HT, pp. 285–294 (2012)
[21] Gayo-Avello, D.: I wanted to predict elections with twitter and all i got was this lousy paper - a balanced survey on election prediction using twitter data. CoRR abs/1204.6441, 1–13 (2012)
[22] Ringsquandl, M., Petkovic, D.: Analyzing political sentiment on twitter. In: AAAI Spring Symposium: Analyzing Microtext, pp. 40–47 (2013)
[23] Unankard, S., Li, X., Sharaf, M.A.: Emerging event detection in social networks with location sensitivity. World Wide Web, 1–25 (2014)
[24] O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: Linking text sentiment to public opinion time series. In: ICWSM, pp. 122–129 (2010)
[25] Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: ICWSM, pp. 178–185 (2010)
[26] Sang, E.T.K., Bos, J.: Predicting the 2011 dutch senate election results with twitter. In: EACL Workshop on Semantic Analysis in Social Media, pp. 53–60 (2012)
[27] Makazhanov, A., Rafiei, D.: Predicting political preference of twitter users. In: ASONAM, pp. 298–305 (2013)
[28] Jungherr, A., Jurgens, P., Schoen, H.: Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, t. o., sander, p. g., & welpe, i. m. "predicting elections with twitter: What 140 characters reveal about political sentiment". Soc. Sci. Comput. Rev. 30(2), 229–234 (2012)