RESEARCH ARTICLE OPEN ACCESS

A Systematic Review of Predicting Elections Based On Social Media Data

Bessy Bijo, D.Raghavi, B.Rishika, G.Gayathri

1. Assistant Professor, Department of Information Technology,
Malla Reddy Engineering College For Women(UGC-Autonomous), Hyderabad, India
Email: bessybijo@gmail.com

2. Department of Information Technology,
Malla Reddy Engineering College For Women(UGC-Autonomous), Hyderabad, India
Email: dyavanapellyraghavi@gmail.com

3. Department of Information Technology,
Malla Reddy Engineering College For Women(UGC-Autonomous), Hyderabad, India
Email: boyinipellyrishika@gmail.com

4. Department of Information Technology,
Malla Reddy Engineering College For Women(UGC-Autonomous), Hyderabad, India
Email:gaddamgayathri289@gmail.com

Abstract:

The way politicians communicate with the electorate and run electoral campaigns was reshaped by the emergence and popularization of contemporary socialmedia (SM), such as Facebook, Twitter, and Instagram social networks (SNs). Due to the inherent capabilities ofSM, such as the large amount of available data accessed in real time, a new research subject has emerged, focusing on using the SM data to predict election outcomes. Main findings include the low success of the most-used approach, namely volume and sentiment analysis on Twitter, and the better results withnew approaches, such as regression methods trained withtraditional polls. Finally, a vision of future research on integrating advances in process definitions, modeling, and evaluation is also discussed, pointing out, among others, the need for better investigating the application ofstate-of the-art machine learning approaches.

I .INTRODUCTION

Social media (SM) has played a central role in politics andelections throughout this decade. We have entered a new era mediated by SM in which politicians conduct permanent campaigns without geographic or time constraints, and additional information about them can be obtained not only by the press but also directly from their profiles on social networks (SNs) and through other people sharing and amplifying their voices on SM. amplifying their voices on SM. In this new scenario, SM isused extensively in electoral campaigns, and an online campaign's success can even decide elections. In practice, recent examples of SM engagement and electoral success include the 2016 U.S. presidential election, when Donald Trump focused his campaign on free-media marketing, and the 2018 Brazilian presidential election, when the candidate with more SM engagement but little exposition on traditional media was elected

elected Moreover, in some way, it is possible to measure how a politician's message is spreading over SM and try to estimate how much attentiona candidate is receiving or how many people are talking about a candidate. Thus, considering a large amount of dataavailable in real time and the low cost of their acquisition, combined with the advances of techniques for processing them, a new research subject has emerged, focusing on using the SM data to predict election outcomes.

Only 2 years after Twitter and Facebook's launch for thegeneral public, studies to predict elections based on the SM data started to be published: Tilton can be considered a preliminary study focused on student elections, published in 2008. In addition, two studies published in 2010 at the same forum, Tumasjan et al. and O'Connor et al. [6], are considered seminal studies regarding predicting political elections based on SM. The former presented an approach based on the volume counting of posts on Twitter (tweets), and the latter was based on the sentiment extracted from those tweets.

One decade after Tumasjan and O'Connor's seminal studies had claimed promising results, several initiatives focused on predicting elections worldwide, such as in Europe, Asia, Latin America, Africa, and the USA, just to cite some.These studies presented a variety of methods, were applied in many different electoral scenarios, used different SNs as an information source, and had different outcomes. Many studies claimed very positive results, others challenged thepredictive power of SM, and even the same study may achieve positive results in one context and negative results in another context [18].
Thus, there is not yet a common perspective on the literature or well-established methods, processes, and tools for predicting election results based on the SM data. Moreover, even the SM context has changed over the years. For example, Facebook surpassed the number of active users of Twitter, and new SN has emerged, such as Instagram.

II RELATED WORK

Contemporary SM systems are new: Facebook and Twitter were launched to the public in 2006, and Instagram emerged in 2010. Only a few days after their launch, SM began to be used in modern political activities and to be considered a source for election prediction. One of the first attempts aimed at predicting election outcomes using data from SM may be attributed to Tilton. In 2008, only two years after Facebook's launch for the general public, he tried to predict election outcomes of a connected society, in this case a university, framed by the following research question: "Could Facebook be used to estimate the results of a student election?" Results showed that his model was able to predict what place the candidates came in 21 out of 27 times in a given election. Probably because it is not related to formal politics scenario, Tilton's study is seldom cited by studies in the area, but we consider it as a very insightful preliminary study in this field Two studies can be considered seminal in predicting political elections with the SM data and are cited by almost all following studies. In 2010, Tumasjan et al. presented a study on the 2009 German federal election. They collected all the tweets with the names of any of the six parties represented in the German parliament, or prominent politicians of these parties, and compared the volume of tweets with the election results. Per their results, they claimed that "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."

Thus, there is not yet a common perspective on the literature or well-established methods, processes, and tools for predicting election results based on the SM data. Moreover,even the SM context has changed over the years. For example, Facebook surpassed the number of active users ofTwitter, and new SN has emerged, such as Instagram.

III. PROPOSED SYSTEM

One of the first attempts aimed at predicting election outcomes using data from SM may be attributed to Tilton [4]. In 2008, only two years after Facebook's launch for the general public, he tried to predict election outcomes of a connected society, in this case a university, framed by the following research question: "Could Facebook be used to estimate the results of a student election?" Results showed that his model was able to predict what place the candidates came in 21 out of 27 times in a given election. Probably because it is not related to formal politics scenario, Tilton's study is

By analyzing 52 studies, 11 regarding election predictions, they identified that main approaches were based on volume, sentiment, and user profiling. In addition, the use of predictive analysis using linear regression was identified, but not on the studies related to the political context. In addition, they verified that 40% of studies that had used sentiment-related variables challenged SM predictive power, i.e., was not successful, and this number increased to 65% in the case of lexicon-based approaches. Finally, they emphasized the lack of predictive analytics evaluation and controversial results of electoral predicting studies. In the same year, Gayo-Avello [30] presented a study that we consider the first review specifically on predicting elections with SM, focused on Twitter. By analyzing ten previous studies from 2010 to 2013, he concluded that "the presumed predictive power regarding electoral prediction has been somewhat exaggerated." Moreover, as in [29], he identified volume and sentiment analysis as main approaches and the need to use more up-to-date methods

he expanded the list of challenges, such as the dependency of arbitrary decisions made by researchers regarding keywords, parties, candidates and selection of the data collection period, and problems related to Twitter, such as demographic and self-selection bias, and bias related to spam, misleading propaganda, and astroturfing.

He ended the study pointing out that regression models may be a future direction. In 2015, studies from Prada and O'Leary presented in general lines the main approaches for predicting using Twitter in many different domains, and briefly described a few studies related to election predictions (2 and 11 studies, respectively). In 2018, Kwak and Cho presented the results of a survey including 69 papers that supported the argument that SM can be used in understanding political agenda, rather than in election forecast. Ultimately, most recent studies presented limited nonsystematic surveys, both analyzing 13 papers, adding some arguments to the original review from Gayo-Avello. Koli et al. argued that prediction using Twitter can have better results in developed countries, due to a higher literacy rate and internet access, than in developing countries.

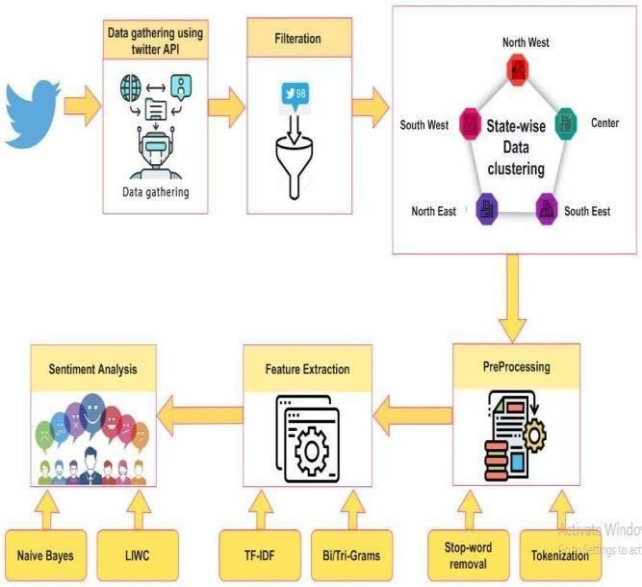
he improved the list of challenges, like the dependency of discretionary selections created through researchers regarding keywords, events, candidates and selection of the facts assortment amount, and troubles related to Twitter, which incorporates demographic and self-selection bias, and bias related to direct mail, deceptive info, and astroturfing. He complete the observe remarking that regression models is a future path. In 2015, analysis from Prada [31] and O'Leary [32] given in standard strains the principle ways for predicting the usage of Twitter in several distinct domain names, and in brief delineate a couple of studies associated with election predictions (2 and eleven studies, respectively). In 2018, Kwak and Cho [33] given the outcomes of a survey that embody sixty nine papers that supported the argument that SM could also be utilized in info political

Communications and electronics systems conference' A universal critical problem has emerged that affects individuals, firms or governments compared with the previous networked systems security.

This paper will look at the ability of online social networksto predict election outcomes of a connected society, in this case a university. Facebook represents a new phenomenon in networking within a university. These network constructsallow for communication to occur rapidly and can influence the opinion of the student body. It is the conglomeration of previous information and communication technologies

is one of most important subjects in

III. SYSTEM ARCHITECTURE



Twitter has become a pervasive tool in election campaigns. Candidates, parties, journalists, and a steadily increasing share of the public are using Twitter to comment on, interact around, and research public reactions to politics. These uses have met with growing scholarly attention. As of now, this research is fragmented, lacks a common body of evidence, and shared approaches to data collection and selection. This article presents the results of a systematic literature review of 127 studies addressing the use of Twitter in election campaigns. In this systematic review, I will discussthe available research with regard to findings on the use of Twitter by parties, candidates, and publics. during election campaigns and during mediated campaign events. Also, I will address prominent research designs and approaches to data collection and selection.

This paper will look at the ability of online social networks to predict election outcomes of aconnected society, in this case a university. Facebook represents a new phenomenon in networking within a university. These network constructs allow for communication to occur rapidly and can influence the opinion of the student body. It is the conglomeration of previous information and communication technologies (ICTs) wrapped up under a simple graphical user interface (GUI) that allows the student body to communicate quickly and has allowed online social networks to

Twitter has become a pervasive tool in election campaigns. Candidates, parties, journalists, and a steadily increasing share of the public are using Twitter to comment on, interact around, and research public reactions to politics. These uses have met with growing scholarly attention. As of now, this research is fragmented, lacks a common body of evidence, and shared approaches to data collection and selection. This article presents the results of a systematic literature review of 127 studies addressing the use of Twitter in election campaigns. In this systematic review, I will discussthe available research with regard to findings on the use of Twitter by parties, candidates, and publics. during election campaigns and during mediated campaign events. Also, I will address prominent research designs and approaches to data collection and selection.

This paper will look at the ability of online social networks to predict election outcomes of a connected society, in this case a university. Facebook represents a new phenomenon in networking within a university. These network constructs allow for communication to occur rapidly and can influence the opinion of the student body. It is the conglomeration of previous information and communication technologies (ICTs) wrapped up under a simple graphical user interface (GUI) that allows the student body to communicate quickly and has allowed online social networks to dominate collegiate culture. Collegiate culture exists in a duality of the real world

The research used a hierarchical linear matrix which was developed forthe work of Auden bush Bryk to develop a model that could answer this question. The final analysis of the matrixshowed it was able to predict what place the candidates came in 21 out of 27 times for all of the candidates in a given election. Inter of predicting the candidate's final percentage of votes received (within half the standard deviation of the Estimated Polling Percentage, which was.072722) during the election 12 out of 27times for all of thecandidates in a given election.

These data show that the research in this area is performed by institutions in all continents, being only two institutions in Africa, and the USA is the country with more institutions involved. Nevertheless, we did not find prominent researchers, research groups, or clusters performing sustainable research in the area. Even the work of was performed in a limited period, mostly in 2011 and 2012, and studies from the University Studies di Milano were limited to 2013 and 2014.

I. IV.LITERATURE SURVEY

The use of social media within modern political activities is a new phenomenon that reshapes election races and the way in which politicians communicate with voters. During the last presidential campaign in Brazil, the elected candidate had almost no time on TV (8 seconds) and very little party support but focused his campaign on social networks. In this context, the objective of this paper is to study the relationship between social media and the electoral performance of candidates running in the 2018 Brazilian presidential election by analyzing how candidates used their social media profiles, assessing how citizens interacted with them, and identifying the correlations between a candidate's performance on social media and votes received. For this, we collected and analyzed the number of followers and all posts from all 13 presidential candidates on the three major social networks, Facebook, Twitter, and Instagram, from January–October 2018. As a result, more than 41,000 posts and 291 million interactions were analyzed, and our findings show that: (i) candidates heavily used social media throughout the year, but focused on engaging words and avoided sensitive topics; (ii) Instagram garnered a higher increase in followers and a higher rate of interactions via posts in comparison to Facebook and Twitter; (iii) there was found no correlation between the number of posts and votes received, with a very small negative correlation with posting about sensitive topics, and a strong correlation between votes and

V.SYSTEM DESIGN

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

This study is carried out to check the economic impact that thesystem will have on the organization. The amount of funds that the company can pour into the research and development of the system is limited. The expenditures must be justified.

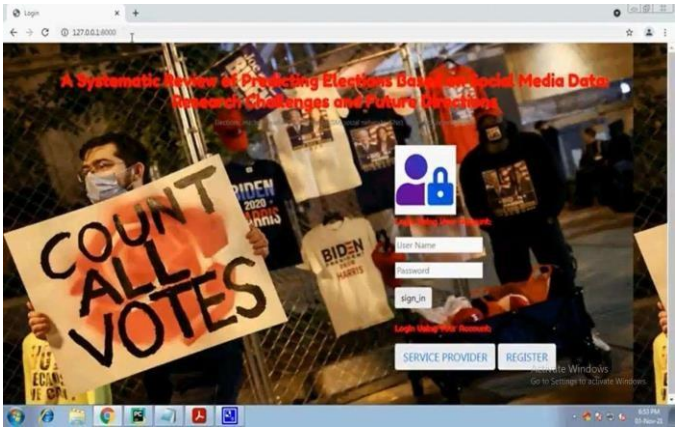
TECHNICAL FEASIBILITY

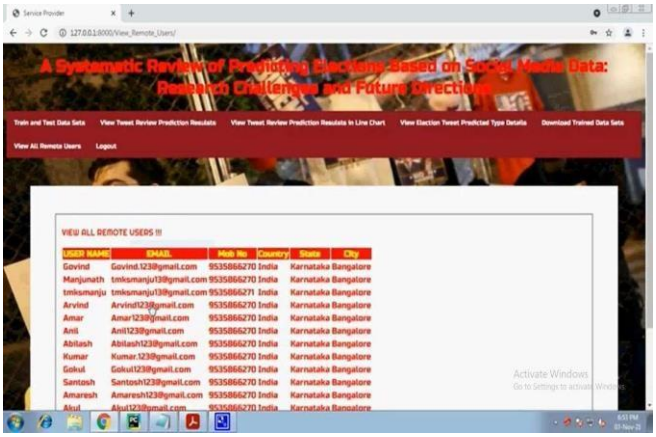
This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand for the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

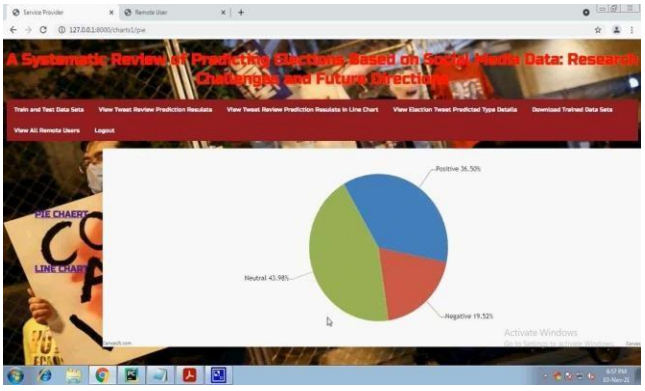
SOCIAL FEASIBILITY

The aspect of the study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

VI.OUTPUTS







js This study collected more than 500 articles, 90 of which were focused on predicting elections based on SM data, investigating, and summarizing how this new research field has evolved since2008. Among these

VII. CONCLUSION

This study collected more than 500 articles, 90 of which were focused on predicting elections based on SM data, investigating, and summarizing how this new research field has evolved since2008. Among these studies, 83 are primary studies aiming at predicting elections and seven are surveys or reviews of past studies.

The results show that the number of publications in this area is increasing and research is spread across 28 countries from all continents. Nevertheless, there cannot yet be found any prominent researchers, research groups, or clusters performing sustainable research in the area. In addition, there was no identification of a common well- known forum for publication on this subject, and results are spread across many forums.

Regarding electoral contexts, most studies were performed in the context of a unique election, which mayimpact the results' validity. In addition, most were related to presidential elections at a national level with few candidates. Moreover, the most studied scenario was the

U.S. presidential scenario, which can impact generalization due to its specificity.

Considering the main models used, we found that most studies used the approach of volume/sentiment analysis only on Twitter, in a variety of data collection approaches. We also found that regression and time series analysis is increasing, using multiple SNs, in addition to some supporting approaches, such as profile or post interactionsand topic analysis.

By combining studies' characteristics and success we found that, despite being the most used approach, volume/sentiment does not present high success rates, which is consistent with the conclusions of previous surveys. Thus, approaches such as regression or based onprofile/post interactions may be better to investigate andimprove; even totally new approaches, such as one based on the statistical physics of complex networks, may be tested. Finally, studies based on Twitter achieved significantly lower success rates than studies based on other SNs, suchas Facebook. Surprisingly, no studies based on Instagramwere found. Moreover, as the main challenges, we identified issues in four areas. Regarding processes, we highlight the lack of well-defined, replicable, and generalizable processes, and the lack of prediction capabilities during the campaign.

VIII. REFERENCE

A. Jungherr, "Twitter use in election campaigns: A systematic literature review," J. Inf. Technol. Politics, vol. 13, no. 1, pp. 72–91, Jan. 2016. P. L. Francia, "Free media and Twitter in the 2016 presidential election: The unconventional of Trump," Social Sci. Computer. Rev., vol. 36, no. 4, pp. 440–455, Aug. 2018.

K. Brito, N. Paula, M. Fernandes, and S. Meira, "Social media and presidential campaigns– preliminary results of the 2018 Brazilian presidential election," in Proc. 20th Annu. Int. Conf. Digit. Government Res., Jun. 2019, pp. 332–341.

S. Tilton, "Virtual polling data: A social network analysis on a student government election," Webology, vol. 5, no. 4, pp. 1–8, 2008.

Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with Twitter: What 140 characters reveal about political sentiment," in Proc. 4th Int. AAAI Conf. Weblogs SocialMedia, 2010, pp. 1–8.

A. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in Proc. 4th Int. AAAI Conf. Weblogs SocialMedia, 2010,pp. 1–8.

E. Sang and J. Bos, "Predicting the 2011 Dutch senate election results with Twitter," in Proc. Workshop Semantic Anal. Social Media, 2012, pp. 53–60.

A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "Every tweet counts? How sentiment analysis of socialmedia can improve our knowledge of citizens' political preferences with an application to Italy andFrance," New Media Soc., vol. 16, no. 2, pp. 340–358, Mar. 2014.

K. Singhal, B. Agrawal, and N. Mittal, "Modeling Indian general elections: Sentiment analysis of political Twitter data," in Information Systems Design and Intelligent Applications (Advances in Intelligent Systems and Computing). New Delhi, India: Springer, 2015.

N. DwiPrasetyo and C. Hauff, "Twitter-based election prediction in the developing world," in Proc. 26th ACM Conf. Hypertext Social Media (HT), 2015, pp. 149–158.

J. A. Ceron-Guzman and E. Leon-Guzman, "A sentiment analysis system of Spanish tweets and its application in Colombia 2014 presidential election," in Proc. IEEE Int. Conf. Big Data Cloud Computer. (BDCloud), Social Computer. Netw. (SocialCom), Sustain. Computer. Communication. (SustainCom) (BDCloud-Social Com-Sustain Com), Oct. 2016, pp. 250–