

Importance of Social Media Analytics During Elections: A Review

P. N. Jain^{1*}, N. V. Alone²

^{1,2}Dept. of Computer Engineering, R. H. Sapat college of Engineering, Management studies and Research, Savitribai Phule Pune University, Nashik, India

*Corresponding Author: pooja.pagaria48@gmail.com, Tel.: +91-7218093959

DOI: <https://doi.org/10.26438/ijcse/v7si14.453458> | Available online at: www.ijcseonline.org

Abstract—The progress of 21st century can barely be anticipated without the indication of the part of social media in it. It wouldn't be overstating to say that social media is ubiquitously present in all spheres of life, be it education, health care, business, disaster management, politics, tourism industry and of course the use of media sharing and entertainment needs no mention. In the wake of all such convenience provided by the social media, it too, does have a darker side to cast. Misuse of social media, the other side of the coin, also needs to be accounted. In the light of this and more so because of the upcoming Lok Sabha elections in India, the authors of this report feel an urge to address the current status of knowledge, the research community possess regarding the use of social media during election. The paper discusses the basics of Social Media Analytics i.e., from its evolution and framework to tool and techniques and also some applications in brief. Finally, several studies on social media analytics during elections have been described. It is sought to contemplate the degree to which the result of an election can be predicted, public opinions be altered or its usefulness in campaigning for an election. Apart from this, the authors also hope that this study will be helpful for other researchers to analyse the social media data and yield productive outcomes that contribute to the development of society, government and the nation.

Keywords—Social Media Analytics, Political science, Elections, Social media.

I. INTRODUCTION

21st century can hardly be thought of without the mention of social media. Our day-to-day lives are largely driven by it. To say the least, even our psychological state, decisions and opinions are influenced on what we encounter on various social media platforms. The number of these platforms increase with increasing involvement of people. Table (1) shows the list of some of the popular social media platforms and the number of active users.

Table 1 List of popular social networking sites and details

Platforms	Active users	Type	Average daily activity
Facebook (2004)	2.32	Social networking	300m photo upload
YouTube (2005)	1.5	Media Sharing	5b video views
Instagram (2010)	1	Media Sharing	95m photos upload
Whatsapp (2009)	0.9	Communication-based	43b messages sent
LinkedIn (2002)	0.6	Social networking	0.17m new account
Reddit (2005)	0.54	Social networking	25m votes

Tumblr (2007)	0.5	Micro blogging	0.24m new blogs
Twitter (2006)	0.3	Social networking	140m tweets
Pinterest (2010)	0.25	Media Sharing	2m visits
Snapchat (2011)	0.2	Communication-based	10b video views
Google+(2011)	0.1	Media Sharing	45% users aged 15–34

Statistics show that the worldwide population is about 7.72 billion where internet users are not less than 4.2 billion. There are about 3.397 billion active social media accounts with 5.54 social media accounts per person. This shows that the data generated by these sites regarding the opinions of public in various affairs like politics, education, healthcare, tourism, crises, terrorism, sports, culture etc. can be extremely helpful for forecasting events, disaster management, emergency response or even spreading awareness. The analysis of this data is generally termed as 'Social Media Analytics' (SMA). As defined by [1], SMA is "an emerging interdisciplinary research field that aims on combining, extending, and adapting methods for analysis of social media data".

The current review paper accumulates all such recent studies from last 10 years, that deal with SMA. More specifically, the focus is on use of SMA during elections. The framework of SMA, its data collection and data analysis procedure have been briefly explained. It is sought to contemplate the degree to which the result of an election can be predicted, public opinions be altered or its usefulness in campaigning for an election. Apart from this, the authors also hope that this study will be helpful for other researchers to analyse the social media data and yield productive outcomes that contribute to the development of society, government and the nation.

A. Evolution and Framework..

The evolution of Social Media Analytics can be contemplated by finding the usage of this term on internet over time and region. Using Google Trends, the worldwide search about the websearch of the term *Social Media Analytics* from 2004 to present can be obtained and this evolution is illustrated in Fig. (1). It can be clearly seen that since 2007, the concept of social media analysis has been increasing sharply and ever since. Also the regional distribution globally reveals that about 32% of the searches have been made in India.

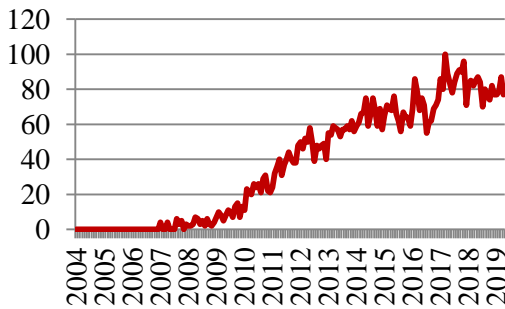


Figure 1. Web search of term *Social Media Analytics* from 2004-present (Using Google trends)

A comprehensive review discussing the big data analysis of social media can be found in [2-4]. To clarify the process, researchers have established frame-works that form a general algorithm for performing social media analytics. The four-step framework are as follows: Discovery, Tracking, Preparation and Analysis as shown in Fig. (2)

B. Tools and Techniques

Over the course of time, many tools, techniques and methods have been developed to address and analyse certain types of data and research questions. Some of the commonly used social media monitoring tools are- Social Mention, Amplified Analytics, Lithium Social Media Monitoring, Trackur, FeedBurner. Analysis of textual data may be done using OpenAmplify, Jodange, GATE, Lexalytics Sentiment Toolkit, AeroText, Attensity, Clarabridge, IBM LanguageWare, SPSS Text Analytics for Surveys, Language

Computer Corporation, STATISTICA Text Miner and WordStat.

Alternatively, data may also be visualised using Apache Hadoop, Amazon Kinesis, SAS Visual Analytics and Tableau. Some of the Keyword based monitoring tools are: Teezir, Hootsuite, Social Mention, Coosto and SocialMediaCheck.

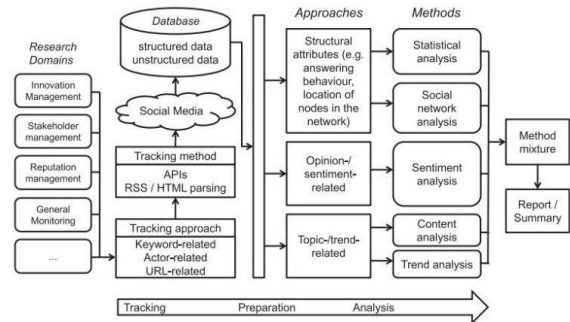


Figure 2. Framework of SMA [4]

Table (2) mentions the tools and methodology used by various researchers for social media analysis to address various research questions of different domains.

Table 2. Social Media Analytics tools and methodologies

Tools / Methodology	References
Sentistrength / Sentiwordnet	[5-8]
Naive Bayes Classifier	[6], [9-11]
Spectral Clustering Algorithm	[12-14]
POS-Tagging	[15-17]
WEKA	[7], [18,19]
R-Square	[20-22]
Latent Dirichlet Allocation Topic Modelling	[9], [23]
LibSVM	[7], [24]
Chi-Squared Attribute	[7], [13], [19], [25]
Regression Approach	[26-29]

C. Data Analysis

Once the collection, storing and cleaning of data is completed, the main task of analyzing the enormous amount of data is to be performed. Tag detection, context-sensitive analysis, community analysis, hashtag analysis, word analysis, feature construction and extraction, casualty analysis, syntactical parsing are some of the type of analysis performed on the data generated by social media data to extract and identify useful information from it. Table (3) demonstrates some commonly implemented type of analysis by various researchers.

Table 3. Type of analysis of SMA

Analysis	References
Sentiment Analysis	[5], [7], [9], [16], [19], [27], [30]

Topic Modelling/ Detection	[12], [17], [19], [25], [31]
Statistical Analysis	[14], [20]
Time Series Analysis	[22], [24]
Regression Analysis	[8,9], [21,22], [27-29]
Temporal Modelling	[8], [23]

II. APPLICATIONS - SMA

It wouldn't be exaggerating to say that social media is ubiquitously present in all spheres of life, be it education, health care, business, disaster management, politics, tourism industry and of course the use of media sharing and entertainment needs no mention. In the wake of all such convenience provided by the social media, it too, does have a darker side to cast. Misuse of social media, the other side of the coin, also needs to be accounted for. Table (4) enlists the applications where social media analytics have been put to use by various researchers from various domains.

Table 4 Applications of SMA

Domain	Description	References
Crises management	Public response during New York's Hurricane Sandy in 2012.	[12], [23]
Tourism	Estimated customers' reviews, consumer satisfaction by regression analysis, comparison of several prominent online review forums	[9], [21,22]
Healthcare	Categorized several online health communities, measure social tension, examined physical activity-based tweets	[6, 7], [10], [19]
Education	Knowledge based social media usage by students	[32 – 35]
Terrorism	Recruitment analysis, use of social media in conflict areas, media mujahedeen', ISIS and SMA	[36 – 45]
Business	Evaluation of stock market short-term performances, impact of major events on stock market and corresponding market volatility	[16, 17], [24, 25], [27]

III. SMA DURING ELECTIONS

Political science is no exception to impact of social media. The amount of data generated on social media concerning politics, might have capacity to predict the polling result, and also influence the voters opinion.

A. Spanish election- 2015

Using twitter data, [46] investigated Spain's general elections in 2015, as the election was seemed to be inclined by the advent of political parties that devoted a major percentage of

their political policy to manage emotions and many infotainment programmes that introduced politicians as guests during the pre-campaign period. The authors structured three extraction principles for data accumulation from twitter: 2 common keywords associated with the election; the username and name of 4 main political parties; name and username of the 4 nominees for prime minister. Additionally, they also dug up tweets in Spanish to completely investigate the content. Authors chose SentiStrength to analyse and adjust the program to fit the corpus. The results suggest an instant positive emotional effect when the contestants' private and cherished matters were argued. Nevertheless, their study did not disclose any noteworthy consequence on the spectators' sense of compassion towards the candidates.

B. Netherlands elections- 2014

The growth of a theoretical framework and a set of rules that adds to quantification of dependable influence of social media campaigns by politicians have been studied by [47] using local municipal elections in the Netherlands in 2014. Data from Facebook, Twitter, YouTube, LinkedIn, and Google were used as Social media resource for analysis. They developed a revised Social Media Indicator 2 (SMI2) which served as an instrument used to report the framework of Social Media usage. During the first pilot test, the comparison of influence scores seemed fairly appropriate using scoring algorithm for Contribution score. Authors related the social media involvement scores to preference votes and discovered positive relationships in several municipalities.

C. Singaporean elections- 2011

By extracting Twitter data during the promotion period of the Singapore General Election of 2011, [48] examined the predictive power of tweeter posts in estimating the election outcomes. Coders developed Twitter crawler using the application programming interface (API) provided by Twitter, a MySQL database and programming language Perl. The crawling process was performed every day to regularly updated the changes after each round. They reported that, at the national level, there is reasonably strong relation between the share of tweets and votes. However, this correlation is considerably feeble at constituency level. Authors further suggest that Twitter info may be more appropriate for conducting macro-level evaluations of political sentimentality rather than for forecasting precise consequences of local elections that are unpredictable and certainly twisted by some persuasive Twitterers.

D. Korean election- 2012

Using several text mining methods of Twitter data regarding the Korean presidential election of 2012, [49] investigated user behaviours, identified and tracked the way in which social issues develop and change. Authors gathered tweets on

Twitter, time stamps and user IDs using Twitter Stream API available in Twitter4J. They stored keywords and mentions, in Redis and My-SQL relational database. Further, they also stored keywords using the Korean Tagging System. For network study and data visualization, authors used the Java Universal Network/Graph Framework which is an open source visualization tool. Additionally, a voltage clustering algorithm was employed for identifying user community. Community detection technique was implemented to the complete system to determine hidden functional or behavioural network units. Gephi was used as a manipulation and visualization software. Authors reported that debatable matters in Twitter are produced, broadcasted, and quenched similar to that by existing media. Further, Twitter users with like political temperaments are disposed to interconnect frequently with their social acquaintances, as indicated by mention-based network analysis. Thus, merely concentrating on the follow/following association can hide real-world judgment behaviour.

E. Croatian election- 2015

The significances of evolution from traditional to social media campaigns and social media's usefulness at activating and altering public opinion in Croatian general election campaign of 2015 has been studied by [50]. Using Facebook's Graph API Explorer, authors recovered all posts and blogs from Croatian CDU and SDP Facebook websites circulated from 25th Sept to 25th Oct 2015. The investigation reported that various political parties apply different election campaign plans on social media to sway voters who, subsequently, retort differently to everyone. The outcome suggests that political communications with positive emotions gives rise to positive reaction from citizens, while neutral contents are probable to appeal for negative criticism and comments.

F. Pakistan election- 2013

The influence of tweets in forecasting the victor of Pakistan general election of 2013 has been examined by [51]. The authors detected appropriate Twitter users, processed those tweets, and structure a predictive model. The prediction was performed using three representative political parties that were considerably posted, namely, Muttahida Qaumi Movement (MQM), Pakistan Muslim League Nawaz (PMLN) and Pakistan Tehreek-e-Insaaf (PTI). Data was collected by means of Twimemachine, that permits to procure the newest tweets of user's. For training purpose, they made use of Rapid Miner tool to examine three predictive models, namely, Support Vector Machine (SVM). Naive Bayes 14 and CHAID decision tree. They labelled prediction labels as Pro and Anti. Anti represents a negative one sentiment and Pro represents a positive one. Authors commented that Twitter by itself cannot be trusted comprehensively for polling forecast due to its evolving Pakistani user-base.

VI. CONCLUSIONS

The usage of Social Media Analytics, predominantly in the last 10 years have been highlighted in this review paper. The framework, mechanism, tools and techniques of SMA have been discussed at the outset. The usage of SMA in various domains like education, healthcare, politics, business, media and tourism, crises management and terrorism have been discussed in brief. The later part of the paper covers, specifically, the usage of SMA during elections and its impact in political sciences. The review indicates that the analysis of social media have been gaining grounds rapidly. Such kind of analysis are highly useful in forecasting, disaster management, emergency response or even spreading awareness. An important research gap that can be highlighted from the study is the need of devising more easier ways of data extraction and data cleaning. Of the huge amount of social media data generated much of it may not only be unstructured but also irrelevant. Hence, data cleaning, sorting and structuring needs additional research.

REFERENCES

- [1] D. Zeng, H. Chen, R. Lusch, and S. Li. "Social media analytics and intelligence". IEEE Intelligent Systems, vol. 25(6), pp. 13–16, 2010.
- [2] N. A. Ghani, S. Hamid, I.A.T. Hashem, and E. Ahmed, "Social media big data analytics: A survey. Computers in Human Behavior", 2018. R. Solanki, Principle of Data Mining, McGraw-Hill Publication, India, pp. 386–398, 1998.
- [3] H. Sebei, M. A. H. Taieb, and M.B. Aouicha, "Review of social media analytics process and big data pipeline", Social Network Analysis and Mining, vol. 8(1):30, 2018.
- [4] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics—challenges in topic discovery, data collection, and data preparation", International journal of information management, vol. 39, pp. 156–168, 2018.
- [5] BK Chae, "Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research". Int. J. Production Econom, vol. 165: pp. 247–259, 2015
- [6] P. Burnap, OF Rana, N Avis, M Williams, W Housley, A Edwards, J Morgan, L Sloan, "Detecting tension in online communities with computational Twitter analysis", Tech. Forecasting Soc. Change, vol. 95, pp. 96–108, 2015
- [7] S Tuarob, CS Tucker, M Salathe, N Ram, "An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages", J. Biomedical Informatics, vol. 49, pp. 255–268, 2014
- [8] Y. Lu, F Wang, R Maciejewski, "Business intelligence from social media: A study from the vast box office challenge". Comput. Graphics Appl., IEEE, vol. 34(5), pp. 58–69, 2014
- [9] Xiang Z, Du Q, Ma Y, Fan W "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism". Tourism Management, vol. 58, pp. 51–65, 2017
- [10] Huh J, Yetisgen-Yildiz M, Pratt W "Text classification for assisting moderators in online health communities". J. Biomedical Informatics, vol. 46(6), pp. 998–1005, 2013
- [11] Abbas A, Ali M, Khan MUS, Khan SU "Personalized healthcare cloud services for disease risk assessment and wellness

- management using social media*". Pervasive Mobile Comput. vol. **28**, pp. **81–99**, 2016
- [12] Huang Y, Dong H, Yesha Y, Zhou S "A scalable system for community discovery in Twitter during Hurricane Sandy." 14th IEEE/ACM Internat. Sympos. Cluster, Cloud Grid Computing (CCGrid) (IEEE, Chicago), pp. **893–899**, 2014
- [13] Hu Y, Gao S, Janowicz K, Yu B, Li W, Prasad S "Extracting and understanding urban areas of interest using geotagged photos." Computers, Environ. Urban Systems, vol. **54**, pp. **240–254**, 2015
- [14] Frias-Martinez V, Frias-Martinez E "Spectral clustering for sensing urban land use using Twitter activity". Engrg. Appl. Artificial Intelligence, vol. **35**, pp. **237–245**, 2014
- [15] Vazquez S, Muñoz-García Ó, Campanella I, Poch M, Fisas B, Bel N, Andreu G, "A classification of user-generated content into consumer decision journey stages". Neural Networks, vol. **58**, pp. **68–81**, 2014
- [16] Vu TT, Chang S, Ha QT, Collier N, "An experiment in integrating sentiment features for tech stock prediction in Twitter". Raghavan S, Ramakrishnan G, eds. Proc. Workshop Inform. Extraction Entity Analytics Soc. Media Data (COLING 2012 Organizing Committee, Mumbai, India), pp. **23–38**, 2012
- [17] Yan Z, Xing M, Zhang D, Ma B, "EXPRS: An extended pagerank method for product feature extraction from online consumer reviews." Inform. Management, vol. **52(7)**, pp. **850–858**, 2015
- [18] Gal-Tzur A, Grant-Muller SM, Kuflik T, Minkov E, Nocera S, Shoor I, "The potential of social media in delivering transport policy goals." Transport Policy, vol. **32**, pp. **115–123**, 2014
- [19] Yoon S, Elhadad N, Bakken S, "A practical approach for content mining of Tweets". Amer. J. Preventive Medicine, vol. **45(1)**, pp. **122–129**, 2013
- [20] Hong S, Nadler D, "Which candidates do the public discuss online in an election campaign? The use of social media by 2012 presidential candidates and its impact on candidate salience." Government Inform. Quart., vol. **29(4)**, pp. **455–461**, 2012
- [21] Ye Q, Law R, Gu B, Chen W, "The influence of user-generated content on traveller behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings." Comput. Human Behav. vol. **27(2)**, pp. **634–639**, 2011
- [22] Zhang Z, Ye Q, Song H, Liu T, "The structure of customer satisfaction with cruise-line services: An empirical investigation based on online word of mouth". Current Issues Tourism, vol. **18(5)**, pp. **450–464**, 2015b
- [23] Chae J, Thom D, Jang Y, Kim S, Ertl T, Ebert DS, "Public behavior response analysis in disaster events utilizing visual analytics of microblog data". Comput. Graphics, vol. **38**, pp. **51–60**, 2014
- [24] Zhang W, Li C, Ye Y, Li W, Ngai EW, "Dynamic business network analysis for correlated stock price movement prediction". Intelligent Systems, IEEE, vol. **30(2)**, pp. **26–33**, 2015a
- [25] He W, Wu H, Yan G, Akula V, Shen J, "A novel social media competitive analytics framework with sentiment benchmarks". Inform. Management, vol. **52(7)**, pp. **801–812**, 2015
- [26] Peterson RD, "To tweet or not to tweet: Exploring the determinants of early adoption of Twitter by House members" in the 111th Congress. Soc. Sci. J., vol. **49(4)**, pp. **430–438**, 2012
- [27] Bollen J, Mao H, Zeng X, "Twitter mood predicts the stock market." J. Comput. Sci., vol. **2(1)**, pp. **1–8**, 2011
- [28] Oh C, Roumani Y, Nwankpa JK, Hu HF, "Beyond likes and Tweets: Consumer engagement behavior and movie box office in social media." Inform. Management, vol. **54(1)**, pp. **25–37**, 2016
- [29] Jin J, Yan X, Li Y, Li Y, "How users adopt healthcare information: An empirical study of an online Q&A community. Internat. J. Medical Informatics, vol. **86**, pp. **91–103**, 2016
- [30] Jang HJ, Sim J, Lee Y, Kwon O, "Deep sentiment analysis: Mining the causality between personality-value-attitude for analyzing business ads in social media. Expert Systems Appl., vol. **40(18)**: **7492–7503**, 2013
- [31] Sun G, Wu Y, Liu S, Peng TQ, Zhu JJ, Liang R, "EvoRiver: Visual analysis of topic competition on social media". IEEE Trans. Visualization Comput. Graphics, vol. **20(12)**, pp. **1753–1762**, 2014
- [32] Kim KS, Sin SCJ, "Use and evaluation of information from social media in the academic context: Analysis of gap between students and librarians". J. Acad. Librarianship vol. **42(1)**, pp. **74–82**, 2016
- [33] Head AJ, Eisenberg MB, "Balancing act: How college students manage technology while in the library during crunch time". Project Information Literacy Research Report, University of Washington Information School, Seattle, 2011
- [34] Junco R, Heiberger G, Loken E, "The effect of Twitter on college student engagement and grades. J. Comput. Assisted Learn. vol. **27(2)**, pp. **119–132**, 2011
- [35] Young CL, "Crowdsourcing the virtual reference interview with Twitter." Reference Librarian, vol. **55(2)**, pp. **172–174**, 2014
- [36] R. M. Medina, "Social network analysis: a case study of the Islamist terrorist network" Security Journal, vol. **27(1)**, pp. **97–121**, 2014.
- [37] S. Borau, and S. F. Wamba, "Social Media, Evolutionary Psychology, and ISIS: A Literature Review and Future Research Directions", In World Conference on Information Systems and Technologies, pp. **143–154**, 2019
- [38] J. Droogan, L. Waldek and R. Blackhall, "Innovation and terror: an analysis of the use of social media by terror-related groups in the Asia Pacific", Journal of Policing, Intelligence and Counter Terrorism, vol. **13:2**, pp. **170–184**, 2018
- [39] A. Fisher, "How jihadist networks maintain a persistent online presence", Perspectives on terrorism, vol. **9(3)**, pp. **3–20**, 2015
- [40] J. Hutchinson, F. Martin and A. Sinpeng, "Chasing ISIS: Network power, distributed ethics and responsible social media research", Internet research ethics for the social Age: New challenges, cases, and contexts, pp. **57–73**, 2017
- [41] Online as the New Frontline: Affect, Gender, and ISIS-Take-Down on Social Media",
- [42] T. E. Nissen, "Terror. com: ISs social media warfare in Syria and Iraq", Contemporary Conflicts, vol. **2(2)**, pp. **2–8**, 2014
- [43] H. N. Teodorescu, "Using analytics and social media for monitoring and mitigation of social disasters", Procedia Engineering, vol. **107**, pp. **325–334**, 2015
- [44] D. Pohl, A. Bouchachia, and H. Hellwagner, "Online indexing and clustering of social media data for emergency management", Neurocomputing, vol. **172**, pp. **168–179**, 2016
- [45] K. Aal, M. Krger, M. Rohde, B. Tadic, and V. Wulf, "Social Media and ICT Usage in Conflicts Areas", In Information Technology for Peace and Security, pp. **383–401**, 2019
- [46] Tomás Baviera, Alvar Peris, and Lorena Cano-Orón. Political candidates in infotainment programmes and their emotional effects on twitter: an analysis of the 2015 spanish general elections pre-campaign season. Contemporary Social Science, vol. **14(1)**: **144–156**, 2019.
- [47] Robin Effing, Jos van Hillegersberg, and Theo Huibers. Social media indicator and local elections in the netherlands: towards a framework for evaluating the influence of twitter, youtube, and facebook. In Social media and local governments, pages 281–298. Springer, 2016.
- [48] Marko Skoric, Nathaniel Poor, Palakorn Achananuparp, Ee-Peng Lim, and Jing Jiang. "Tweets and votes: A study of the 2011 singapore general election." In 2012 45th Hawaii international conference on system sciences, pp. **2583–2591**. IEEE, 2012.
- [49] Min Song, Meen Chul Kim, and Yoo Kyung Jeong. "Analyzing the political landscape of 2012 korean presidential election in twitter." IEEE Intelligent Systems, vol. **29(2)**, pp. **18–26**, 2014.

- [50] Marina Bađić Babac and Vedran Podobnik. “What social media activities reveal about election results? the use of facebook during the 2015 general election campaign in croatia”. *Information Technology & People*, vol. 31(2), pp. 327–347, 2018.
- [51] Tariq Mahmood, Tasmiyah Iqbal, Farnaz Amin, Wajeeta Lohanna, and Atika Mustafa. “Mining twitter big data to predict 2013 pakistan election winner”. In *INMIC*, pp. 49–54. IEEE, 2013.

Authors Profile

Ms. Pooja Naginkumar Jain

BE – Pune University, 2013

ME – Pune University, pursuing,

Research area: Imaging analysis, social media analysis, social networking analysis, smart homes etc.



Prof. Nilesh V. Alone

BE – Pune University

ME – Pune University

Research area: Internet of Things, Wireless sensor networks, Machine learning, and Social media analysis
