Predictive Modeling of Public Opinion for Karnataka Elections using Twitter Data Analysis

Annie Syrien
Department of Computer Applications, Banglore University
Bengalore, India
annie.syrien@sju.edu.in

Dr. M. Hanumanthappa
Department of Computer Applications, Banglore University
Bengalore, India
hanu6572@gmail.com

***Abstract*—The Karnataka state assembly elections have brought a new perspective to the political landscape, with diverse candidates representing various parties vying for the people's choice. This research work provides new insights into the predictive aspects of the Karnataka elections. Through the use of Deep Learning models applied to data collected from Twitter, the sentiments and past experiences of the ruling party were analyzed to predict the future elected party. The focus is on the public's sentiments with regards to the hopes for democratic policies in the future. The data was collected through the Twitter API and transformed into a well-structured format for training the model with a 70:30 ratio. The performance metric of the model is tabulated. Data exhibits mixture of opinions from the people of Karnataka towards the ruling party. While some tweets expressed confidence in the change of government formation in the future, others expressed concerns on religion and anti-national statements in a sarcastic manner. The LSTM model produced results with an accuracy of 87%.**

***Keywords*—LSTM, Twitter, Karnataka Elections, Deep Learning models, Machine Learning**

I. Introduction

Sentiment Analysis is a paramount application of the machine learning [1], With the proliferation of Social Networking Sites(SNS) and microblogs [2], like Twitter, also the advancement of 4G and 5G technologies, data augmentation has become increasingly necessary [3, 4] . In particular, Twitter has become a popular microblog where users can express their opinions in short-form content [5]. In 2017, Twitter enhanced the character restriction of tweets from 140 to 280, allowing users to be more expressive [6]. With over 400 million users worldwide, Twitter is a widely-used platform where people search for articles, news, and trends [7]. In the United States alone, there are 37 million active Twitter users, while India has 24.45 million Twitter users, according to statistics from the website Backlinko.

The grail of the research grind is to analyze the sentiments expressed on Twitter regarding the 2023 Karnataka Assembly elections and classify them. This entails a thorough examination of the language utilized in tweets, necessitating the adoption of advanced Natural Language Processing(NLP) and Deep Learning(DL) techniques to decipher the intricate nuances and implications of public sentiment as expressed. [8, 9] Illustrates the implication of social media in political communication.

The thrust of elections is to upload the democratic values which encompasses a equality, freedom, and participation in government decisions [10]. The objective of voting involves the citizens participation in political, to promote democracy, to promote transparency, to hold accountability [11]. The current elections in Karnataka are the sixteenth assembly elections, with a total of 224 seats. A political party needs to secure a minimum of 113 seats to form a government in the state. The party with the majority in the assembly forms the government [12].

The research work grails to accurately predict the sentiments of people in Karnataka using a deep learning model. The data is possessed from Twitter's Application Programming Interface (API), preprocessed, and vectorized. The deep learning model, specifically an LSTM, is used to detect sentiments in the tweets. Finally, performance metrics are computed to evaluate the model's effectiveness. The introduction of the research paper outlines the objective and motivation of the study. The related work section presents recent studies in the field and compares their results. The following section details the methodologies used, including the LSTM model, and presents the results. The conclusion condenses the overall analysis of the research work.

The motivation for conducting research on the Karnataka elections is multifaceted. On one hand, it involves leveraging deep learning models to gain insight into the political landscape and voting patterns in the state. Additionally, there is a desire to understand the potential outcomes of the election and what they may mean for the future of Karnataka. Moreover, there is a sense of satisfaction in raising awareness about the electoral process and encouraging civic engagement among the public.

II. Related Work

The literature review extensively enlists the diverse works carried out in the multifaceted field of computer science and political science. Anuradha et al [13], evinces a framework for elections through social media, the focal work is to prognosticate the referendum via Twitter data. The research work spotlights the political parties are increasingly using Social Media(SM) platforms like Twitter to reach out to voters and promote their campaign messages. The study uses machine learning(ML) algorithms to analyze Twitter data during the 2014, Indian general elections and predicts the election outcomes.

© The Author(s) 2024
A. K. Visvam Devadoss et al. (eds.), *Proceedings of the 6th International Conference on Intelligent Computing (ICIC-6 2023)*, Advances in Computer Science Research 107,
https://doi.org/10.2991/978-94-6463-250-7_30

In 2018, the influence of social media on the US presidential election results demonstrated the significant impact of these platforms, which have attracted a vast audience worldwide for a variety of purposes, from connecting with friends to garnering likes from other users. Through her research, Manisha Madhava et al [8] has played an influential role in identifying the role of SM in the election of political ruling parties. Specifically, the work provides insights into the impact of SM on the Karnataka assembly elections of 2018 and offers a broader perspective on the trajectory of Indian elections.

The grail of the study, is to investigate the factors that determine voting behavior in India. The work of Patibandla Anil et al [14] comprises of two segments. Ideology and hawkishness. first is based on various authors construe the behavior of voters based on his ideology. The second is on empirical occurrence and patterns of voting in current political frolic. The paper emphasizes the definition of "voting behavior", which encompasses individual psychological patterns, political action, and institutional factors, such as communication media, and their influence on elections.

Neena Talwar Kanungo [15] foresights on India's smudging of digital elections in the year 2015. The composition evinces prominent political parties such as NCP, BJP, BSP, INC, CPI and AAP during maneuver in 2014. The data is excavator from Twitter and Facebook over the period of 3 months. The composition exhibits the political parties strategies with regard to the voters mindset and invades them. SM has been the easy access place for many political parties for easy canvass. It is also anticipated to be a cutting edged for 16th assembly elections.

Usha M Rodrigues [16] culminates the limitation on political canvassing on internet. Some of the issues pertaining the number of people accessing the internet and accessibility of the remote area persons are discussed in the research. The research work mainly takes the 2014 elections for case history. Though the only the significant number of users could possibly use the internet, the tweets posted by political parties made significant difference in readers mind. It provided multi- dimension platform for the parties, where the people of different class of society are reached in various ways. The tweets in a continues basis engages the voters and also brings an awareness on the political dimensions for users. the number of first-time voters and levels of internet accessibility.

The researchers has taken the various dimensions for SM analysis including regional based, lingual based, voters based, sentiment based and so on. Widodo Budiharto et al [17], explored SNS like twitter and taken the Indonesia presidential election for demonstration of same.

An algorithm was created to differentiate the important data and the model was trained with the polarity of the tweets. R language was used for election prediction model. The authors very happy with the tweets results as it synched with the factual out-turns of the elections. results.

Ankita Sharma et al [18] worked on sentiment analysis on election data. The process involved the acquisition, preprocessing, and analysis of tweets, followed by sentiment analysis using different approaches. This study collected tweets from January to March 2019 to gauge public opinion regarding the general elections in India. Two candidates, Candidate-1 and Candidate-2, were analyzed. The study found that Candidate-1 was more popular and well-liked compared to Candidate-2. These results were in similar to the existent aggregate of the May 2019.

Kawaljeet Kaur Kapoor et al [13] reconnoiters the success of the BJP party in the prime ministerial elections in 2014. According to the proposed credo, the amalgamation of SM and election canvassing helped the jog miles of era. It was very easy key for BJP to meet the people. The research work, stresses on fast penetration of election campaigning was successful with the aid of SM. The various illustrations are given in how the contesting party made good absorption of the opportunity of using the SM effectively.

The study [20] compared public opinions from the 2016 and 2020 US Presidential Elections, analyzing tweets collected through the API, pre-processing the data, and applying the Naive Bayes Classifier to extract sentiment. We identified outliers, analyzed swing states, and cross-validated the election results with social media sentiment. Our findings showed that social media sentiment often aligned with the election outcomes, with pre and post-election sentiment analysis highlighting shifts in outliers.

Kokil Jaidka [21] contributed for predicting elections through comparative study on divergent electoral model. The work comprises of three method and three country. The tweets on India, Malaysia and Pakistan are acquired. The different mechanisms for sentiment classification was explored. Merging of method and sentiment together for SA. Similarly Sentiment and volume, and emotion and SNS, yielded better results in terms of voter interest and outcome based discussions in regional part. It was noteworthy, that the out-turns of the countries like India and Pakistan were directed. Traditional polling strategies were not sufficient to penetrate all the people. The tweets demonstrating different opinion on various parties have been influential factor for fellow users.

As the elections neared the number of tweets were gradually increased leading to user checks on tweets. Since the data collected for various countries had the regional limitation, multilingual resources also turned to be a challenging. Hence the some regional parties sentiments classification was not feasible.

III. Methodology

The focus of the study is on the upcoming Karnataka Assembly Elections in 2023. To brace the data for analysis, a numerous preprocessing steps were executed. These included the abolition of stop words and the application of Regular Expressions(RE) to clean the data. Spell checking was performed using the textblob library, and stemming was carried out using the Porter stemmer algorithm. Once the corpus was built, one-hot encoding was applied to represent the text, followed by padding to ensure uniform sentence length. Word embedding was then used to identify the positional relationships between words in the text. The classification task was performed using a Bi-directional LSTM model with a sigmoid activation function. To minimize the loss value, a dropout ratio of 0.3 was set. The methodology of the model is represented in figure 1. The twitter API gave provision for data collection with an assistance of R programming.

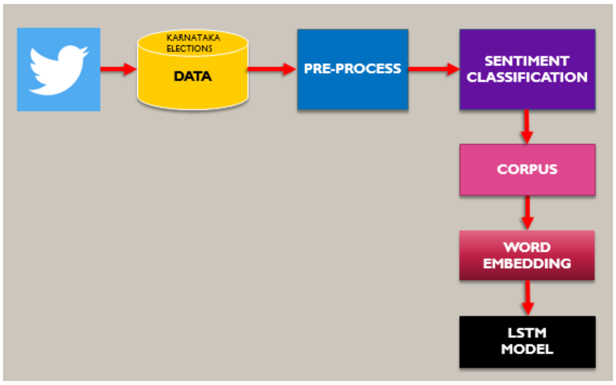


Fig. 1. Methodology

The elections process involves registration of voters, the nomination and campaigning of candidates, the voting process itself (which may take various forms such as paper ballots, electronic voting, or mail-in ballots), and the counting and verification of votes to determine the winner [22]. The election process may be regulated by government bodies or independent organizations to ensure fairness and transparency [23]. The figure 2 illustrates the election process.

Fig. 2. Election process

A. Data

The dataset was built from the Twitter API by generating Authorization tokens using the R tool. The dataset included various fields such as tweet text, favorites, reply information, tweet creation date, and more. We collected over 44,400+ tweets related to Elections collected using appropriate keywords, which were then preprocessed for further analysis. The figure 3 gives the glimpse of tweets. The figure 4 shows the most negative tweet and figure 5 most positive tweet from the extracted dataset.

.

	slno	text
1	1	RT @CvoterSurvey: 21 -22 April Tracker Result , Opinion Poll Of Karnataka Elections 2023 TOTAL SEAT - 224 <U+25AA><U+F...
2	2	@GurmeSuresh May you win the next coming Karnataka State elections. All the very best Suresh Ji <U+0001F389><U+2728>
3	3	RT @SSPBLR: @rinpoet Don't worry He has the fullest support of our enemy countries. Once the Karnataka elections get over,...
4	4	RT @arungogs: @prettypadmaja @IndiaToday @siddaramaiah When Siddaramaiah was the CM in 2017 it was under his eye "...
5	5	RT @BSBommai: Who would be the next CM, only BJP high command can tell on May 13. Since it's under my leadership, I am...
6	6	Karnataka Elections: Shettar, who left BJP to join Congress, welcomes Rahul Gandhi at the airport https://t.co/i8mzJ13pQs
7	7	RT @annamalai_k: Today, in Bengaluru, spoke at the @IndiaToday Round Table about #DMKFiles, the TN FM's audio tape leak,...
8	8	RT @TimesNow: 'BJP's Victory Is Certain': CM Bommai Kicks Off Campaign For Karnataka Polls https://t.co/anLoNSdwH8
9	9	#KarnatakaElections2023 : #CPI extends support to #Congress to 'safeguard democratic and constitutional values' https://t.co...
10	10	RT @kcvenugopalmp: We are one team, together we will conquer! Karnataka Congress is on the way to deliver a historic vict...
11	11	'BJP's Victory Is Certain': CM Bommai Kicks Off Campaign For Karnataka Polls https://t.co/anLoNSdwH8
12	12	RT @INCIndia: Congress President Shri @kharge chaired the first meeting of AICC Observers for the upcoming Karnataka elec...
13	13	RT @ramupatil_TNIE: #KarnatakaPollsWithTNIE: #PoliticalRadar: Battle lines drawn: @INCIndia keen to keep up the momentu...
14	14	RT @shaandelhite: I gurantee you 200% that Congress will touch 146 seats in the upcoming assembly elections of Karnataka. ...
15	15	RT @INCIndia: It only goes to prove that the BJP and corruption go hand in hand. The people of Karnataka will give a fitting r...
16	16	RT @Deshpande__: Rajdeep repeatedly pressing Khargeji to become CM of Karnataka; ignoring his age and seniority. Very dis...

Showing 1 to 16 of 44,410 entries, 10 total columns

Fig. 3. Data View

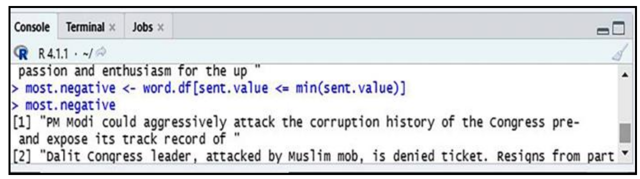


Fig. 4. Sample negative tweet

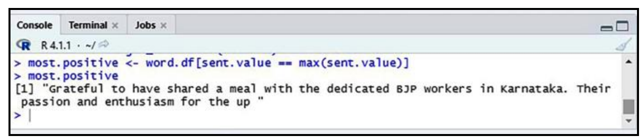


Fig. 5. Sample positive tweet

B. LSTM

LSTM model is an idiosyncratic type of Recurrent Neural Network(RNN) that can effectively manoeuvres long-term dependencies and learn from past memories. Unlike traditional RNNs, LSTMs incorporate a memory cell and 3 gates like forget gate, input gate, and output gate. These components work together to help the model retain relevant information while discarding irrelevant information, making them particularly useful for applications involving sequential data. The memory cell is responsible for storing information over time, while the forget gate allows the model to selectively discard information that is no longer relevant.

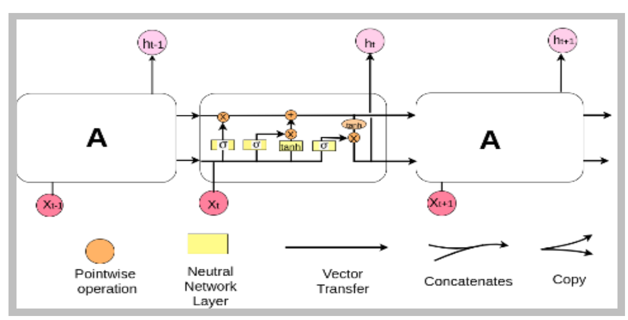


Fig. 6. LSTM model

The LSTM model is represented in the figure 6. As a neural network it arrogates a sequence of characters as input and classifies it into either positive or negative opinion. The data flows through the network via arrow lines, the out-turn of particular node imply as the input for the other. The network architecture is composed of rectangular boxes that represent the various neural network layers, and the merging lines indicate the concatenation of data. Forking lines indicate the copying of multiple pieces of data to different locations within the network. The circles in the network represent point-wise operations that are used to manage and manipulate memory within the network.

Cell state: The cell state in the LSTM model leverages a memory cell to selectively retain or discard information from the input text based on its context. The memory cell is manipulated through a set of point-wise operations that are depicted in the figure 7. Specifically, the yellow circle with an x mark indicates the point-wise operation that is used to forget or discard some of the information in the memory cell, while the yellow circle with a plus mark indicates the point-wise operation that is used to add new information to the memory cell. These operations allow the LSTM model to dynamically update its memory and selectively retain the relevant information for downstream analysis.

Forget Cell: In the LSTM model, the forget gate is a critical component that selectively discards data out of memory cell. The forget gate is the first gate in the LSTM structure and plays a prime role in maintaining the relevant information for downstream analysis. The computation of f_t involves the concatenation of two sets of weights, w_i and w_{h-1}, which are represented by the variable W_f. The output of the forget gate, f_t, is then used to update the memory cell, allowing the LSTM model to selectively retain or discard information based on its relevance to the current context.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

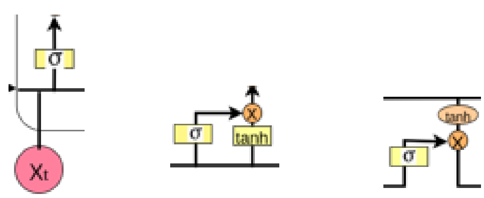


Fig. 7. forget gate, input gate and output gate

Input gate: Through the input gate, information is sent to the cell state and it is adjacent to input gate. It is in charge for selectively retain or discard data.

Output gate: The output gate is a crucial component of the LSTM model that determines which information should be made visible to the next layer of the neural network. This gate is located adjacent to the input gate and works in tandem with the forget gate and input gate.

The gates in the LSTM model utilize a sigmoid activation function, which produces a binary output that decides which information should be retained and which information should be discarded. If the binary value is low, the information will not be passed through the cell state, while if the binary value is high, the information will be passed through the cell state. The LSTM model is particularly effective at processing sequential data and has demonstrated strong performance in tasks that require the model to selectively remember or forget certain pieces of information.

IV. Results and Discussions:

The results measured in terms of recall, precision, accuracy, F1-score , macro mediocre, weighted mediocre and support. In the case of classification tasks, such as predicting the outcome of an election, there are several metrics that can be used to measure the quality of the predictions made by the model.

Precision is a metric that articulates the proportion of exactly predicted positive occurrence out of all instances predicted as positive. The model records the precision of 90%.

Recall is a metric that articulates the proportion of exactly predicted positive occurrence out of all actual positive instances. The model records the recall of 92%.

F1-score is a benchmark for balancing the precision and recall. It is the harmonic mean between precision and recall. An elevated F1-score elevates that the models is ability to make accurate positive predictions while also correctly identifying a high proportion of positive instances. The model records F1-score of 91%.

Accuracy is a criterion that appraises the proportion of suitably predicted instances out of all precedents. It calculates the proportion of true positives and true negatives over the total instances. High accuracy indicates that the model is making accurate predictions overall. The accuracy of 87% is recorded.

Macro mediocre and weighted mediocre are metrics used to evaluate performance across multiple classes in a multi-class classification problem. Macro mediocre calculates the average of the metric across all classes, while weighted mediocre calculates the average weighted by the count of instances in apiece class. The values are recorded similar to the accuracy.

Finally, support is a metric that indicates the number of instances in each class. It can be useful to identify imbalanced classes, where the occurrence of specimens in sole class is consequentially sizeable or minuscule than the others.

Each metric provides a different perspective on the quality of the predictions, and together they can give a comprehensive view of how well the model is performing. The results are illustrated in the figure 10.

A. *Word Cloud*

A word cloud(WC) is a graphical portrayal of a text data set, where the intensity of each locution is proportional to prevalence in the given text. The words are typically displayed in a random arrangement, and can be color-coded to indicate various attributes such as sentiment or topic. Word clouds provide an easy-to-understand visual summary of the most frequently occurring words in a text, making it a popular tool for quickly identifying key themes and patterns in large datasets. The figure 8 describes the WC of 2023 elections.

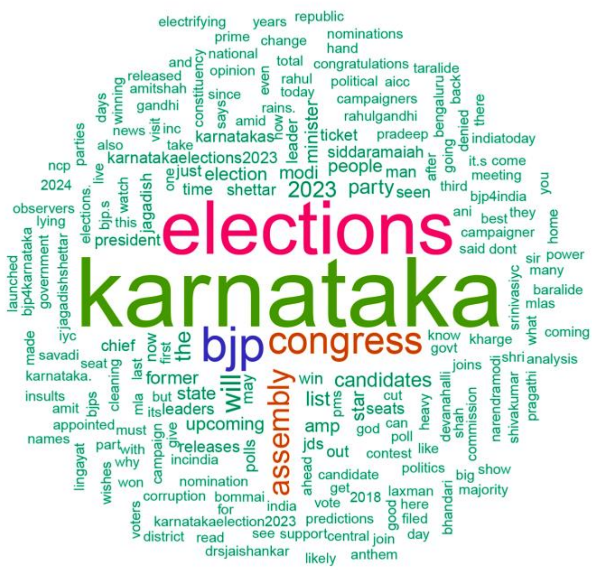


Fig. 8. Word Cloud

B. Epoch

In DL/ML models the datasets will pass through sufficient training in passes. Each of this pass will be refered as epochs. The various occurence of epochs during training can increase the performance. The figure 9 shows the screenshot of different epoch cycles.

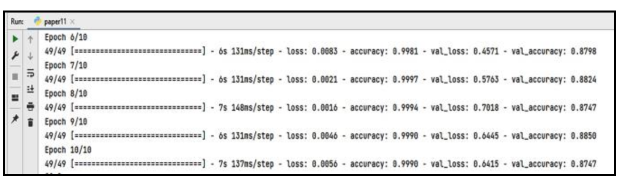
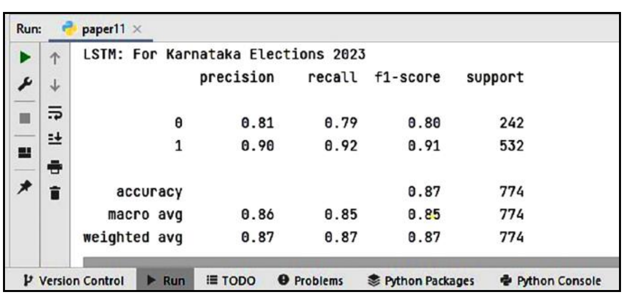


Fig. 9. Epoch cycles



Run: paper11

LSTM: For Karnataka Elections 2023

	precision	recall	f1-score	support
0	0.81	0.79	0.80	242
1	0.90	0.92	0.91	532
accuracy			0.87	774
macro avg	0.86	0.85	0.85	774
weighted avg	0.87	0.87	0.87	774

Version Control Run TODO Problems Python Packages Python Console

Fig. 10. Results of LSTM model

V. CONCLUSION:

The research paper explores the use of the LSTM deep learning technique to analyze tweets related to the upcoming Karnataka Assembly Elections in 2023. The LSTM model is applied to sequential data to determine public opinion on the elections. The model exhibits strong performance on text data, thanks in part to the sigmoid activation function used in the LSTM gates to regulate the data retrieval process. The LSTM model has potential applications in various fields, including text analysis, translation, and speech processing, where accuracy is crucial. The ramification of the research designates that the LSTM model outperforms existing approaches in terms of producing better results. Furthermore, the study found that sarcastic tweets expressing dissatisfaction with the ruling party were prevalent, and that people in Karnataka had mixed opinions about the upcoming elections. Future work could include analyzing the sentiments of users in regional languages and clustering individual tweets based on user accounts.

REFERENCES

[1] M. Choy, M. L. F. Cheong, M. N. Laik and K. P. Shung, "A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction," arXiv preprint arXiv:1108.5520, 2011.

[2] J. Ramteke, S. Shah, D. Godhia and A. Shaikh, "Election result prediction using Twitter sentiment analysis," in 2016 international conference on inventive computation technologies (ICICT), 2016.

[3] M. Ibrahim, O. Abdillah, A. F. Wicaksono and M. Adriani, "Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation," in 2015 IEEE international conference on data mining workshop (ICDMW), 2015.

[4] T. A. Neyazi and R. Schroeder, "Was the 2019 Indian election won by digital media?," The Communication Review, vol. 24, no. 2, pp. 87-106, 2021.

[5] O. Almatrafi, S. Parack and B. Chavan, "Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014," in Proceedings of the 9th international conference on ubiquitous information management and communication, 2015.

[6] N. Narasimhamurthy, "Use and rise of social media as election campaign medium in India," International Journal of Interdisciplinary and Multidisciplinary Studies, vol. 1, no. 8, pp. 202-209, 2014.

[7] K. Carney, "The Effect of Social Media on Voters: Experimental Evidence from an Indian Election," Job Market Paper, pp. 1-44, 2022.

[8] M. Madhava, "Changing Dynamics of Political Electioneering in India: Use of New Social Media in Karnataka Assembly Elections 2018," in Political Communication in Contemporary India, Routledge India, 2023, pp. 52-62.

[9] K. Singhal, B. Agrawal and N. Mittal, "Modeling Indian general elections: sentiment analysis of political Twitter data," in Information Systems Design and Intelligent Applications: Proceedings of Second International Conference INDIA 2015, Volume 1, 2015.

[10] P. Sharma and T.-S. Moh, "Prediction of Indian election using sentiment analysis on Hindi Twitter," in 2016 IEEE international conference on big data (big data), 2016.

[11] P. Keshav, B. Himani and S. Taha, "Indian social media politics: new era of election war," Вестник Российского университета дружбы народов. Серия: Литературоведение, журналистика, vol. 25, no. 1, pp. 184-192, 2020.

[12] M. Madhava, "5 Changing Dynamics of Political Electioneering in India," Political Communication in Contemporary India: Locating Democracy and Governance, 2022.

[13] G. Anuradha, N. R. Moparthi and S. Namballa, "Predicting elections over Twitter: a campaign strategies of political parties using machine learning algorithms," International Journal of Advanced Intelligence Paradigms, vol. 21, no. 3-4, pp. 305-320, 2022.

[14] P. Anil, "Determinants Of Voting Behavior In Indian Politics: A Special Reference To The Recent Assembly Elections In Karnataka," Editorial Board, vol. 7, no. 11, p. 28, 2018.

[15] N. T. Kanungo, "India's digital poll battle: Political parties and social media in the 16th Lok Sabha elections," Studies in Indian Politics, vol. 3, no. 2, pp. 212-228, 2015.

[16] U. M. Rodrigues, "The media, the social media and the elections," India election, pp. 114-124, 2014.

[17] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," Journal of Big data, vol. 5, no. 1, pp. 1-10, 2018.

[18] A. Sharma and U. Ghose, "Sentimental analysis of twitter data with respect to general elections in India," Procedia Computer Science, vol. 173, pp. 325-334, 2020.

[19] K. K. Kapoor and Y. K. Dwivedi, "Metamorphosis of Indian electoral campaigns: Modi's social media experiment," International Journal of Indian Culture and Business Management, vol. 11, no. 4, pp. 496-516, 2015.

[20] H. N. Chaudhry, Y. Javed, F. Kulsoom, Z. Mehmood, Z. I. Khan, U. Shoaib and S. H. Janjua, "Sentiment analysis of before and after elections: Twitter data of us election 2020," Electronics, vol. 10, no. 17, p. 2082, 2021.

[21] K. Jaidka, S. Ahmed, M. Skoric and M. Hilbert, "Predicting elections from social media: a three-country, three-method comparative study," Asian Journal of Communication, vol. 29, no. 3, pp. 252-273, 2019.

[22] M. Ranganathan, "Indian elections, 2014: Commercial media pushes social media into focus," Asia Pacific Media Educator, vol. 24, no. 1, pp. 23-38, 2014.

[23] H. Wang, D. Can, A. Kazemzadeh, F. Bar and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," in Proceedings of the ACL 2012 system demonstrations, 2012.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.