# Prediction of Political Election Results in India using Sentiment Analysis on Twitter Reviews: Survey

**Sayali Jori**
Symbiosis Institute of Technology, Pune, Maharashtra, India

*Abstract*— Election is conducted to view the public opinion, where group of people choose the candidate by using votes, many methods are used to predict result. Many agencies and media companies conduct pre poll survey and expert views to predict result of election. The twitter data is to predict outcome of election by collecting twitter data and analyze it to predict the outcome of the election by analyzing sentiment of twitter data about the candidates.

***Key words:*** Sentiment Analysis, Natural Language Processing, Training Data

## I. INTRODUCTION

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. Sentiment analysis aims to detect opinions expressed regarding a given subject or topic from text. With the rapid growth of social media platforms such as micro blogging services, social networking sites and short messaging services, people increasingly share their views and opinions online. As such, sentiment analysis has attracted much attention since opinions or sentiments detected from text are potentially useful for downstream applications including recommender systems, social network analysis, market forecasting and the prediction of political topics.

In product reviews, it is observed that the distribution of polarity ratings over reviews written by distinct users on different products are often skewed in the real world. Opinion mining refers to the application of text mining, computational linguistics, and NLP, to identify or classify the opinion expressed in text message is either positive or negative. As such, incorporating user and product information would be helpful for the task of sentiment classification of reviews.

The sentiment analysis can be performed either at document level, at sentence level, or at aspect level. At the document level, opinion mining consists of identifying the overall sentiment polarity as expressed in a review.

### A. *Sentiment Analysis*

Sentiment Analysis (SA) elucidates users whether information or opinion regarding a certain product is positive, negative or neutral. Sentiment basically refers to any opinion or a feeling expressed by someone. Various organizations use this analysis to understand users' opinion for their products. For example, a particular e-commerce website can utilize sentiment analysis to discern if their products are being liked by the customers or not. The reviews for the products can be generalized into positive or negative as well as neutral categories. SA can be simply put as "What other people think?"

The terms views, belief, sentiment and opinion can be defined as follows:

1) Opinion- A conclusion open to dispute
2) View- A subjective opinion
3) Belief- Deliberate acceptance and intellectual assent
4) Sentiment- opinion representing someone's feelings

### B. *Different Classes of Sentiment Analysis*

Sentiments can be classified into three class's .i.e. positive, negative and neutral sentiments.

*1) Positive Sentiments*

These are the good words about the target in consideration. If the positive sentiments are increased, it is referred to be good. In case of product reviews, if the positive reviews about the product are more, it is bought by many customers.

*2) Negative Sentiments*

These are the bad words about the target in consideration. If the negative sentiments are increased, it is discarded from the preference list. In case of product reviews, if the negative reviews about the product are more, no one intend to buy it.

*3) Neutral Sentiments*

These are neither good nor bad words about the target. Hence it is neither preferred nor neglected.

### C. *Sentiment Analysis Using Different techniques*

There are basically three techniques to perform Sentiment Analysis.

1) SA using machine learning.
2) SA using lexicon based techniques
3) SA using the above two techniques combined together.

*1) Machine learning technique involves both supervised and unsupervised learning.*

1) Unsupervised Learning is based on just inputs, without any mention of targets. It just relies on clustering.
2) Supervised Learning defines pre-specified targets which should be achieved, along with the inputs. Data set are trained to achieve significant outputs when encountered during decision-making.

*2) Lexicon-Based Approaches*

Lexicon based method assigns positive or negative polarity based on the sentiment of each word and then a dictionary is created. We can use a combining function, for example, sum or average to find out the general sentiment of a document.

### D. *Word Wise Sentiment Analysis*

In word based approach the criteria of selecting Twitter sentiments with the presence of words that express sentiments such as good, bad, excellent, trouble and etc. From these words it is possible to infer the sentiment present in the text. These words are used to determine the Twitter sentiments and must be created for each sentiment that is positive or negative according to the applications.

*1) Using Senti Word Dictionaries*

SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.

 

a) SentiWordmeans

– A general thought, feeling, or sense.
– Feelings, especially tender feelings, as apart from reason or judgment.
– Gentle or tender feelings, sometimes of a weak or foolish kind.

b) Using WordNet Dictionaries

WordNet is a combination of dictionary and thesaurus. It groups English words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members.

c) Using NLP and Sentiment Dictionaries-NLP-

Natural language processing is ontology-assisted way of programming in terms of natural language sentences.

Natural language processing (NLP), is a branch of artificial intelligence that concerned with automated interpretation and generation of human language.

d) POS Tagging Wise

Part-Of-Speech (POS) detects if the word token is noun, verb, and adjective. The word is assigned in accordance with its syntactic functions. In English the main parts of speech are noun, pronoun, adjective, determiner, verb, preposition, adverb, conjunction, and interjection.

e) NER Detection Wise

Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names. Classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

f) Emoticon Wise Sentiment Analysis

In this approach the criterion to select Twitter sentiments for classification is the presence of at least one emoticon. Based on emoticons used in the message it is possible to infer the sentiment of the text

The Emoticon data set is created by collecting status and comments with positive ':)' and negative ':(' emoticons. In this approach messages are classified based on positive and negative emotions. Using Unicode's we can do sentiment analysis.

| Emoticons | Feeling | Sentiments |
|---|---|---|
| :) :-) | Happy | Positive |
| :( :-( | Sad | Negative |
| :D :-D | Very Happy! | Positive |
| D: D= | Very Sad | Negative |
| *_* *.* *-* | Fascinated | Positive |
| D:< D: D8 | Horror, disgust, sadness | Negative |
| xD XD | Laughing, big grin | Positive |
| :\| =\| :-\| | Straight face no expression | Neutral |

Table: Sample of Emoticons Used

g) Hybrid Sentiment Analysis

Hybrid sentiment analysis is a combination of all the above 3 techniques such as Word wise sentiment analysis, Emoticon wise sentiment analysis, and SVM algorithm wise sentiment analysis. It gives result in the enhancement and increases the system performance and accuracy.

## II. Literature Review

"Aldo Hernández", et.al introduces sentiment analysis method on Twitter content to predict future attacks on the web. The method is based on the daily gathering of tweets from two sets of users; the individuals who utilize the platform as a method for expression for views on relevant issues, and the individuals who utilize it to present contents identified with security attacks in the web. Predicting attacks is an imperative task that considers what actions ought to be taken if the assault is latent. The daily information is converted into data that can be broke down statistically to predict whether there is a plausibility of an assault. The last is finished by investigating the aggregate sentiment of users and groups of hacking activists in response to a global event. The goal is to predict the response of specific groups involved in hacking activism when the sentiment is sufficiently negative among various Twitter users. For two contextual analyses, it is demonstrated that having coefficients of determination greater than 44.34% and 99.2% can figure out whether a significant increase in the percentage of negative opinions is identified with attacks.

"Mauro Dragoni, Fondazione Bruno Kessler" et al the brief information about the tool that are for enabling the building of multi-domain sentiment model that gives the linguistic overlaps between domains for inferring document polarity of each domain. It makes the use of a deep learning architecture using distributed vectors to represent words. It can be used for calculating the polarity of a given text by using domain-specific information. Dranziera protocol is used for evaluating the performance of the system.

"Manju Venugopalan" , et.al describe building up a half and half model for sentiment classification that explores the tweet specific features and uses domain independent and domain specific lexicons to offer a domain oriented approach to analyze sentiment of shoppers regarding different smart phone brands. The analyses have demonstrated that the results enhance by around 2 points on an average over the unigram baseline. The SVM accuracy has improved in the range 1.5 to 3.5 and J48 could provide an accuracy improvement ranging from 1.5 to 4 points across domains. The improved lexicon which have adapted polarities learning the domain and the tweet specific features extracted have added to the improvement in classification accuracies.

"LakshmishKaushik, AbhijeetSangwan and John H. L. Hansen" et al LakshmishKaushik et al. focuses on a new method for recognizing sentiment in audio using keyword spotting (KWS) for sentiment detection with the use of audio obtained from videos in youtube.com and UT-Opinion corpus. The system uses the iterative methodology to automatically extract sentiment behavior keywords from text. The Maximum Entropy based approach is used to enlarge the sentiment classifier.

"Ming Hao", et.al introduces novel techniques three novel time based visual sentiment analysis techniques to explore high volume of Twitter data. These techniques are: (1) topic-based sentiment analysis that extracts, maps, and measures customer opinions; (2) stream analysis identifying interesting tweets depending on density, negativity, and impact attributes; and (3) pixel cell based sentiment timetables and high density geo maps that visualize substantial volumes of data in a single view. These techniques were connected to a variety of twitter data, (e.g., movies, amusement parks, and hotels) to demonstrate their distribution and patterns, and to recognize influential opinions. A visual analysis of Twitter time series was displayed, to explore equivalent Twitter data streams.

"Yoonjung Choi, JanyceWiebe, and Rada Mihalcea" et al discuss a knowledge-based +/-effect coarse-grained sense disambiguation method to convey the opinion towards +/-effect of the word sense depending on the surrounding context that positively or negatively affect entities. The system uses the selection preferences which are modeled using Latent Dirichlet Allocation (LDA).The system uses WordNet information is used to determine whether an instance of a word in the corpus is being used with a +effect, -effect, or Null sense.

"RuiXia ,Jie Jiang, and Huihui He" et al introduces A distantly supervised lifelong learning approach, for large-scale social media sentiment analysis is proposed in for addressing distant supervision work in terms of continuously increasing and constantly changing topics. The continuous sentiment learning in social media can learned the knowledge from past tasks, and continuously update the knowledge as new tasks appear. The system can adapt any single-task sentiment learning algorithms to the scene of lifelong sentiment learning. The lifelong sentiment classifier is evaluated on nine benchmark datasets.

"Rincy Jose", et.al focuses on Natural Language (NLP) approach to enhance sentiment classification by adding semantics in feature vectors and thereby using ensemble methods for classification. Generally, bag-of-words approach has been used for mining sentiments online. In this approach, individual words are considered instead of complete sentences. Traditional machine learning algorithms such as Support vector Machines, Naive Bayes' and Maximum entropy etc. are commonly used to solve the classification problems there is a certain level of bias toward a particular class using above techniques. Therefore, Natural Language (NLP) based approach has been used to enhance the sentiment classification. Conducted experiments have shown that semantics based feature vector gives 3-5% better results than the above mentioned bag of words approach.

"Fuhai Chen, Rongrong Ji, Jinsong Su, Donglin Cao, and Yue Gao", et al discuss a Weakly Supervised Multi-modal Deep Learning (WS-MDL) scheme towards robust and scalable sentiment prediction. The system learns convolutional neural networks iteratively and selectively from "weak" emoticon labels, which mostly contains noise. A probabilistic graphical model is explained to filter out the emoticon labels serve as noisy labels to be used and evaluated during training that simultaneously learn discriminative multi-modal descriptors and infer the confidence of label noise.

"Jyoti Ramteke; Samarth Shah: Darshan Godhia: Aadil Shaikh", proposed two stage framework which can be used to create a training data from the mined Twitter data without compromising on features and contextual relevance. Finally, they proposed a scalable machine learning model to predict the election results using our two stage framework.

"Gayatri P. Wani", et al introduces the general user tweets from the election point of view. Here the System will study the user view of Indian election. Based on the users tweets system analyses if there exist a pattern between the tweets and to analyze and draw meaningful inferences from the collection of these tweets collected over certain period; the proposed system identify the feasibility of development of a classification model to identify the political orientation of the twitter users based on the tweet content and other user based features. There are voting advice applications (VAAs) are online tools which are popularly used in deciding which party/candidate to vote for during an election in countries like Greece, Cyprus but still there in India there is no such application which focus on this .

"Pritee Salunkhe", et al discuss that researchers dealing with utilizing twitter to monitor people reactions in political activities like debates and campaigns. On the basis of that prediction an election can be made. Analysis of the prediction of election results using messages of either political parties or politician. The use of tweet content considered as a valid indicator of political sentiment. Sentiment analysis is used for analysis as well as predicting the emotions from the text patterns regarding political issues, products, entity, election etc. Sentiment analysis consists analyzing texts to extract information. Basic sentiment analysis allows to determine or measuring the polarity (negative or positive) of sentiment.

"Abhishek Bhola", analyzed the complete dataset to find interesting patterns in it and also to verify if the trivial things were also evident in the data collected. We found that the activity on Twitter peaked during important events related to elections. It was evident from our data that the political behavior of the politicians affected their followers count and thus popularity on Twitter. Yet another aim of our work was to find an efficient way to classify the political orientation of the users on Twitter. To accomplish this task, we used four different techniques: two were based on the content of the tweets made by the user, one on the user based features and another one based on community detection algorithm on the retweet and user mention networks. We found that the community detection algorithm worked best with an efficiency of more than 80%.

"Ali Hasan" et al introduces growth in the area of opinion mining and sentiment analysis has been rapid and aims to explore the opinions or text present on different platforms of social media through machine-learning techniques with sentiment, subjectivity analysis or polarity calculations. Despite the use of various machine-learning techniques and tools for sentiment analysis during elections, there is a dire need for a state-of-the-art approach. To deal with these challenges, the contribution of this paper includes the adoption of a hybrid approach that involves a sentiment analyzer that includes machine learning. Moreover, this paper also provides a comparison of techniques of sentiment analysis in the analysis of political views by applying supervised machine-learning algorithms such as Naïve Bayes and support vector machines (SVM).

"Akshat Bakliwa" et al discusses a series of 3-class sentiment classification experiments on a set of 2,624 tweets produced during the run-up to the Irish General Elections in February 2011. Even though tweets that have been labelled as sarcastic have been omitted from this set, it still represents a difficult test set and the highest accuracy we achieve is

61.6% using supervised learning and a feature set consisting of subjectivity-lexicon-based scores, Twitter specific features and the top 1,000 most discriminative words. This is superior to various naive unsupervised approaches which use subjectivity lexicons to compute an overall sentiment score for a pair.

## III. Advantages

1) Adjust market strategy.
2) Improve customer service.
3) It provides audience insight.
4) Social media popularity and election results.

## IV. Applications

1) Political exit poll prediction
2) Best visiting places recommendation
3) Movies comparisons and calculate ratings.

## V. Conclusion

This work begins with transforming unstructured information into meaningful lexicons after extracting the Twitter's contents. All of the meaningful lexicons are stored in a database after manual identifications are carried out. With sentiment analysis, emotions are classified into happy (positive), unhappy (negative) and emotionless. The results are displayed by giving the percentage of sentiment categories so that it can be concluded that a selected Twitter post get positive or negative responses based on all comments received from users. As a case study, an issue on an examination results is posted and results of students' responses are determined. Th is study is significant of enabling the stakeholders such as administrators and businessmen to monitor any discussion issue for enhancing their services.

## References

[1] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub,"Arabic sentiment analysis: Lexicon-based and corpus-based," in AppliedElectrical Engineering and Computing Technologies (AEECT), 2013IEEE Jordan Conference on. IEEE, 2013, pp. 1–6.
[2] R. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis inArabic tweets," in Information and communication systems (icics),2014 5th international conference on. IEEE, 2014, pp. 1–6.
[3] Assiri, A. Emam, and H. Aldossari, "Arabic sentiment analysis: Asurvey," vol. 1, no. 6, 2015, pp. 75–85.
[4] M. Abdul-Mageed, M. T. Diab, and M. Korayem, "Subjectivity andsentiment analysis of modern standard Arabic," in Proceedings of the49th Annual Meeting of the Association for Computational Linguistics:Human Language Technologies: short papers-Volume 2. Associationfor Computational Linguistics, 2011, pp. 587–591.
[5] M. Nabil, M. A. Aly, and A. F. Atiya, "Astd: Arabic sentiment tweetsdataset." in EMNLP, 2015, pp. 2515–2519.
[6] M. Abdul-Mageed and M. T. Diab, "Sana: A large scale multi-genre,multi-dialect lexicon for Arabic subjectivity and sentiment analysis." InLREC, 2014, pp. 1162–1169.
[7] Pang and L. Lee, "A sentimental education: Sentiment analysis usingsubjectivity summarization based on minimum cuts," in Proceedings ofthe 42nd annual meeting on Association for Computational Linguistics.Association for Computational Linguistics, 2004, p. 271.
[8] M. Lan, S.-Y. Sung, H.-B.Low, and C.-L. Tan, "A comparative studyon term weighting schemes for text categorization," in Neural Networks,2005. IJCNN'05.Proceedings.2005 IEEE International Joint Conference on, vol. 1. IEEE, 2005, pp. 546–551.
[9] Jyoti Ramteke: Samarth Shah: Darshan Godhia: Aadil Shaikh "Election Result Prediction Using Twitter sentiment Analysis" IEEE 2016.
[10] Mitali Desai: Mayuri A. Mehta "Techniques for sentiment analysis of Twitter data: A comprehensive survey" IEEE 2017.
[11] Nugroho Dwi Prasetyo "Tweet-Based Election Prediction" TUDeltft 20104
[12] Bholane Savita Dattu, Prof.Deipali V. Gore "A Survey on Sentiment Analysis on Twitter Data Using Different Techniques" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5358-5362.
[13] Parul Sharma and Teng-Sheng Moh "Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter" 2016 IEEE International Conference on Big Data.
[14] Brandon Joyce: Jing Deng "Sentiment analysis of tweets for the 2016 US presidential election" IEEE 2017.
[15] Ali Hasan , Sana Moin , Ahmad Karim and Shahaboddin Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts"Math. Comput. Appl. 2018, 23, 11; doi:10.3390/mca23010011www.mdpi.com/journal/mca.
[16] Gayatri P. Wani and Nilesh V. Alone "Analysis of Indian Election using Twitter" International Journal of Computer Applications (0975 – 8887) Volume 121 – No.22, July 2015.
[17] Pritee Salunkhe, Avinash Surnar, Sunil Sonawane "A Review: Prediction of Election Using Twitter Sentiment Analysis" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET-2017).
[18] Akshat Bakliwal , Jennifer Foster , Jennifer van der Puil, Ron O'Brien, Lamia Tounsi and Mark Hughes "Sentiment Analysis of Political Tweets: Towards an Accurate Classifier" 2013 Association for Computational Linguistics.
[19] Abhishek Bhola "Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General Elections2014.
[20] Aparup Khatua, Apalak Khatua, Kuntal Ghosh, Nabendu Chaki "Can #Twitter_Trends Predict Election Results? Evidence from 2014 Indian General Election" 48th Hawaii International Conference on System Sciences. IEEE 2015.