

Forecasting Indian Elections Setup Based on Examining Sentiment Analysis

Amandeep Kaur¹

Research Scholar, Ph.D.(Comp. Appl.)
Guru Kashi University, Talwandi Sabo
Punjab, India.

Dr. Rajinder Singh²

Assistant Professor, UCCA
Guru Kashi University, Talwandi Sabo
Punjab, India.

Abstract: Sentiment analysis has become the mainstream source for obtaining ideas, building strategies, and predicting future outcomes irrespective of business, sector, and field. The paper elaborates the steps or methods to be adopted for conducting a procedural sentiment analysis. The paper also discusses the predicted outcome of Indian elections to be held in 2019. The paper also discusses about classifying the keywords in any post or tweet as positive, negative, or neutral sentiment.

Keywords – Elections, negative words, positive words, sentiments, sentiment analysis, social media.

I. INTRODUCTION

Sentiment analysis refers to the term that mentions the use of computational linguistics, text analysis, and natural language processing in order to determine the approach of a person towards any particular topic. It enables one to judge whether a text is communicating sentiments which are negative, positive, or neutral. Sentiment analysis is broadly used today to mine information from the content present on the Internet like comments, reviews, social media, tweets, and blogs. This can be performed using different techniques like machine learning, NLP, and statistics. Organizations make use of this mined information to recognize new opportunities and target demographics. The popular example of use of sentiment analysis is of Obama administration which used sentiment analysis to predict public response [1, 2].

Sentiments can be classified into two broad categories mentioned as under [3, 4].

- **Subjectivity/Objectivity Identification:** This entails categorizing a sentence or its fragment into one of the two categories: subjective or objective. However conducting this analysis is a challenging task. The major challenge is that the meaning of the word depends on its context.
- **Feature/Aspect-based identification:** It allows the determination of different sentiments in accordance with different aspects of an entity. It allows nuanced impression of opinions and feelings.

There is a wide use of sentiment analysis in today's world. With tremendous increase in social media sites like Twitter and Facebook, the popularity of blogs, increase in reviews and ratings has contributed to growing interest in sentiment analysis. Today consumers can easily

share their experiences with any particular product or service online very much ease. It is because of this that online opinions have become a valuable asset for organizations busy in identifying new opportunities, cultivate their reputations, and successfully launch their products into the market. Organizations are analyzing the power of sentiment analysis to mine information to better understand the customer's interests and buying patterns. Different applications of sentiment analysis are mentioned as under [5, 6].

Social Media Monitoring: Companies today utilize automated sentiment analysis based on word lists, with every word assigned a pre-defined value. The text value is calculated based on the words contained in it. There are number of different uses of such analysis. For instance, a restaurant is involved in social media monitoring to determine what are the customers opinion regarding their menu, find out whether people liked their food and service, and how the overall experience of the people at the restaurant. The good news is that the accuracy associated with sentiment analysis in reference with social media monitoring is continuously getting better. Companies have gained the accuracy of over 75 percent with automated sentiment analysis. Another advantage of sentiment analysis is that in case where there are limited number of people and few of them have a continuous negative thinking about any particular service or product, the owner can reach such people personally and handle the situation accordingly to change their negative perception to positive one.



Fig. 1 The figure shows the impact of social media monitoring

Public Relations: Sentiment analysis also finds its application in building and refining organizations public relation strategies. By forming proper strategies, which require proper mining of available data, the company can individually target its customers and know their buying patterns, interests, and likes and dislikes.

Marketing: Companies are making use of customer generated product reviews on social media. For example, Samsung is interested in knowing the consumers views about its newly launched Galaxy phone. The concerned analysts evaluate the comments posted by customers on major e-commerce sites like Amazon and Flipkart. On examining the tone of the comments written by the customers, Samsung can gain insight of what customers feels about their new launch [7].

Data Mining: Sentiment analysis can even allow one to perform appropriate data mining and gain competitive advantage over one's competitors. For example, a brand could easily track social media mentions or mentions of competitors in other places across the web, and analyze how consumers feel about the competitors and their products. This is better method to gain insight about one's competitor weaknesses and strength and plan own strategies accordingly.

Political Analysis: Studies of sentiment analysis of tweets and microblogs have shown that such analysis can accurately indicate political sentiment. Due to the advent of data explosion, there is an emerging need to collect, monitor, analyze, summarize, and visualize politically relevant information from social media and other online sources. In order to analyze such an open and unstructured domain, it is essential to employ generic and robust methods to procure, integrate, store, and analyze large amounts of text-based data from multiple sources and contexts. The Indian political system remains largely unanalyzed, mainly owing to the unavailability of structured textual data openly accessible by the public. There are still a number of different challenges when it comes to efficiently conducting sentiment analysis. First and foremost, people tend to express their opinions in complex ways, which can make it difficult to identify a clear sentiment. Furthermore, in an opinion text, lexical content alone can make it difficult to determine the opinion being expressed, while verbal devises such as irony and sarcasm can also make it more difficult to clearly identify sentiment. However, as technology continues to improve, it will become easier and easier to overcome these challenges [8, 9, 10].

II. CLASSIFICATION OF WORDS IN DIFFERENT CATEGORIES

Language is extremely difficult to interpret especially in cases where one is reading the text which has been written by such people who are from social group different from one's group. There are even cases where people speak your language but because of their accent, it becomes difficult to judge that whether they are talking about good experience or bad one. For instance, if one analyze a movie and use the word "bomb" in reference to a movie, a mature person would take it as negative or bad movie but if the same word is used by someone belonging to youth movement, it would be interpreted as a positive comment. This proves that language is used in different ways and is constantly evolving [11].

There are multiple approaches to choose from on social media while analyzing sentiments. The practice evolves parsing text, extracting strings, and applying rules. This approach seems sensible in simple situations. But in complex situations, manually coding text-extractions

quickly becomes no longer possible from the viewpoint of code conservation, especially for performance optimization.

In large-scale extraction in Hadoop, the rule based grammatical approaches are computationally expensive. The more complicated are the rules, as in case of sentiment extraction, the greater processing is needed. The statistical approach is become more common for sentiment analysis. The classification oriented machine learning models like Apache Mahout are available which avoids the use of complex rules. The intensive training is required to train the models with negative and positive sentiment. The more is the training, better are the results [12].

The primary job to be performed while mining sentiments is to classify different words related to emotions, feelings, and sentiments in to two broad categories, i.e., positive words and negative words mentioned as under in Fig. 2 and Fig. 3.

└

Positive words

Abound abunds abundance abundant accessible acclaim acclaimed acclamation accolade accolades accommodative amaze amazed amazement amazes amazing amazingly ambitious ambitiously ameliorate backbone balanced bargain bauteous beautiful beautifully beautify beauty beckon beckoned beckoning beckons believable beloved benefactor beneficent beneficial beneficially beneficiary benefit benefits benevolence benevolent benefits best best-known best-performing best-selling better better-known better-than-expected beautifully blameless cajole calm calming calmness capability capable capably captivate captivating carefree cashback cashbacks catchy celebrate celebrated celebration celebratory champ champion charisma charismatic charitable charm charming charmingly chaste cheaper cheapest cheer cheerful cheery cherish cherished cherub chic chivalrous chivalry civility civilize clarity classic classy clean cleaner cleanest cleanliness cleanly clear clear-cut cleared clearer clearly clears daring daringly darling dashing dauntless dawn dazzle dazzled dazzling dead-cheap dead-on decency decent decisive decisiveness dedicated eager eagerly eagerness earnest earnestly earnestness ease eased eases easier easiest easiness easing easy easy-to-use easygoing ebullience ebullient ebulliently economical fabulous fabulously facilitate fair fairly fairness faith faithful faithfully faithfulness fame famed famous famously fancier fascinating fancy fanfare fans fantastic fantastically gaiety gaily gain gained gainful gainfully gaining gains gallant gallantly galore geekier geeky gem gems generosity generous generously hail halcyon hale hallmark hallmarks hallowed handier handily hands-down handsome handsomely handy happier happily happiness happy hard-working hardier

Fig. 2 The figure shows the prominent positive words used on social media

Negative words

Abnormal abolish abominable abominably abominate abomination abort
aborted aborts abrade abrasive abrupt abruptly abscond absence
absentminded absentee absurd absurdity absurdly absurdness abuse
abused abuses abusive abysmal abysmally abyss accidental accost
accursed accusation accusations accuse accuses accusing accusingly
acerbate acerbic acerbically ache ached aches achy aching acrid acridly
acridness acrimonious acrimoniously acrimony adamant brutalities
brutality brutalize brutalizing brutally brute brutish buckle bug bugging
buggy bugs bulkier bulkiness bulky bulkiness bull conspirator
conspiratorial conspire consternation contagious contaminate
contaminated contaminates contaminating contamination contempt
contemptible degenerate degenerately degeneration degradation degrade
degrading degradingly dehumanization dehumanize deign deject dejected
dejectedly dejection delay delayed delaying delays delinquency
delinquent delirious delirium delude deluded deluge delusion delusional
delusions demean demeaning demise demolish demolisher demon
demonic demonize demonized demonizes demonizing demoralize
demoralizing demoralizingly denial denied denies denigrate denounce
dense exasperating exasperatingly exasperation excessive excessively
exclusion excoriate excruciating excruciatingly excuse excuses execrate
exhaust exhausted exhaustion exhausts exorbitant exhort exile exorbitant
exorbitantly expel expensive expire expired explode fear fearful fearfully
fears fearsome feckless feeble feebly feebleminded feign feint fell felon
felonious ferociously ferocity fetid fever gripe gripes grisly gritty gross
grossly grotesque grouch grouchy groundless grouse growl grudge
grudges

Fig. 3 The figure shows the prominent negative words used on social media

III. SENTIMENT ANALYSIS IN PREDICTING INDIAN ELECTION SCENARIO 2019

On the basis of the collected data from various authentic sources on performing sentiment analysis from multiple sources like Facebook, Twitter, and various blogs, it can be predicted that the chances of BJP to retain power in 2019 elections is quite bright. But it can also be equally predicted that the magical majority mark may be difficult for saffron party to achieve.

The biggest hindrance to Modi's ambition for second term is anti-incumbency and unity of opposition parties. In 2004, the opinion poll indicated that Vajpayee led NDA would continue with its second term because of the popularity gained through BJP's "India Shining" campaign. But India has different plans and elected Manmohan Singh led government to rule India for next decade. In 2009 election, the Congress returned with even better mandate as compared to 2004. The Congress led government won 261 seats out of 543. But in 2014, the BJP surprised everyone by winning 282 seats. But the story would not be same in 2019. BJP may lose more than 50 seats in its eight key states. The figure below depicts the detailed prediction for elections 2019 [13, 14].

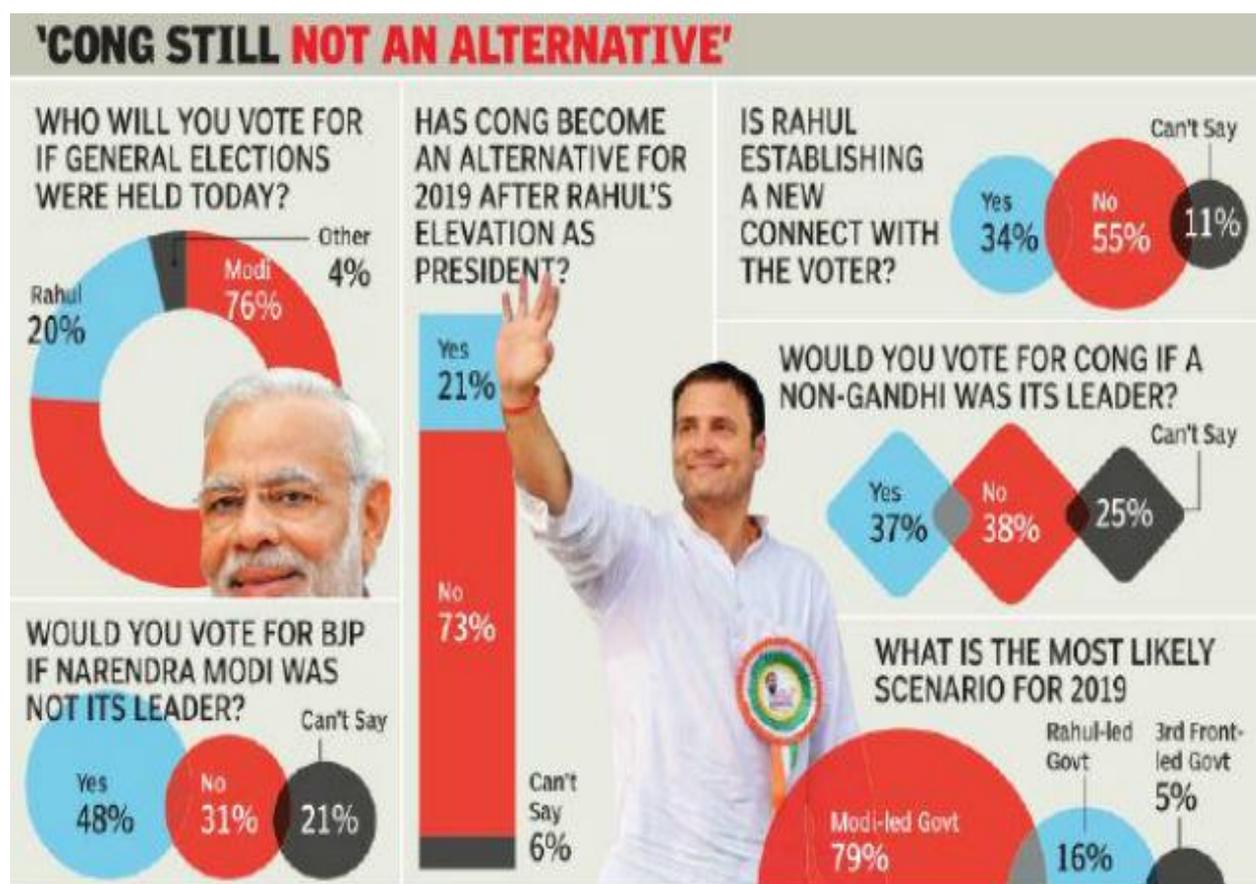


Fig. 4 The figure shows the predicted result for upcoming Indian Elections 2019

IV. STEPS TO BE PERFORMED FOR CONDUCTING SENTIMENT ANALYSIS

Below mentioned are the steps to be performed for conducting sentiment analysis [15, 16].

Data collection

The data should be collected from different authentic origins and should be send to a resting location like HDFS for later analysis.

Labelling the data

This is business relevant part of the process. There is a need to recognize words that are applicable to one's business to come up with the kind of dictionary and to attribute to words and expressions a polarity (positive, neutral/negative) or a note (from 0 to 10, 5 being neutral). Hadoop embeds customizable catalogues and dictionary tables to help one in performing the task [17].

Running the analytics

The next step is to score the sentiment of the tweets which is done by performing comparison between numbers of positive words to the number of negative words. As the data is present in HDFS, the tables can be created in Hive.

Training and adapting to update the model

The analytics tools which concentrate only on positive or negative words are totally misleading if they miss on significant context. Intentional misspellings, typos, jargon, and emoticons are some of the obstacles in this task.

Computers also don't understand sarcasm and irony and as a general rule are yet to develop a sense of humour. Too many of these and one will lose accuracy. It's probably best to try to address this point by fine-tuning one's model [18, 19].

Get insights!

When done, simply run some interactive queries in Hive to refine the data and enjoy visualization of data via a BI tool. Depending on one's business, Hadoop will certainly enable one to take urgent marketing decisions and actions. This is just one of many ways to collect and analyse social data using Hadoop and there are myriad other options open to be explored-it's all about what is right for you!

V. CONCLUSION

The paper discussed the methods to perform sentiment analysis from different social media sources. The paper also elaborated on the predictions of the upcoming elections in 2019. The political parties can plan out their strategies based on the sentiment analysis performed on the data from different authentic sources.

REFERENCES

1. B. Agarwal, N. Mittal, "Prominent Feature Extraction for Review Analysis: An Empirical Study", In Journal of Experimental and theoretical Artificial Intelligence, 2014, DOI: 10.1080/0952813X.2014.977830.

2. S Bandyopadhyay and K Mallick, "A New Path Based Hybrid Measure for Gene Ontology Similarity", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol.11, no. 1, pp. 116-127, Jan.-Feb. 2014, doi:10.1109/TCBB.2013.149
3. Jagdev, G. (2018). Sentiment Analysis and its Impact in Modeling Election Scenario. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 5(2), pp.22-27. <http://dx.doi.org/10.20431/2349-4859.0502004>.
4. Mariana Romanyshyn(2013). Rule-Based Sentiment Analysis of Ukrainian Re-views. International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 4, July 2013
5. Gagandeep Jagdev et al., "A Comparative study of Conventional Data Mining Algorithms against Map-Reduce Algorithm", in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN (O) – 2319-8354, ISSN (P) – 2319-8346, Volume – 06, Issue – 05, May 2017.
6. Gagandeep Jagdev et al., "Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data", in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550, Volume – 05, Issue – 05, May 2017.
7. Jose, R. and Chooralil, V. S. (2015). Prediction of election result by enhanced sentiment analysis on twitter data using word sense disambiguation. In Control Communication & Computing India (ICCC), 2015 International Conference on, pages 638–641. IEEE.
8. Amandeep Kaur, & Dr. Gagandeep Jagdev (2017). Exploring Application of Big Data in Elections – From Data to Action, International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 4(4), pp.64-71, DOI: <http://dx.doi.org/10.20431/2349-4859.0404008>.
9. Gagandeep Jagdev et al., "Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data", in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550, Volume – 05, Issue – 05, May 2017.
10. Padmaja, S., Fatima, S. S., and Bandu, S. (2013). Analysis of sentiment on newspaper quotations: A preliminary experiment. In Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on, pages 1–5. IEEE.
11. Gagandeep Jagdev et al., "A Study of Clustering and Classification Techniques involved in Data Mining", in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550, Volume – 05, Issue – 05, May 2017.
12. Gagandeep Jagdev et al., "Analyzing and Filtering Big Data concerned with elections via Hadoop Framework" in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN (O): 2319-8354, ISSN (P): 2319-8346, Volume No. 6, Issue No, 4, April 2017.

13. Sharma, A. and Dey, S. (2012). A comparative study of feature selection and machine learning techniques for sentiment analysis. In Proceedings of the 2012 ACM Research in Applied Computation Symposium, pages 1–7. ACM.
14. Gagandeep Jagdev et al., “Comparing Conventional Data Mining Algorithms with Hadoop based Map-Reduce Algorithm considering elections perspective”, in International Journal of Innovative Research in Science and Engineering (IJIRSE), ISSN: 2454-9665 (O), ISSN: 2455-0663(P), Volume – 3, Issue – 3, March 2017.
15. Gagandeep Jagdev et al., “Big Data Proposes an Innovative concept for contesting elections in Indian Subcontinent” in International Journal of Scientific and Technical Advancements (IJSTA), ISSN-2454-1532, 2015.
16. Gagandeep Jagdev et al., “Excavating Big Data associated to Indian Elections Scenario via Apache Hadoop” in International Journal of Advanced Research in Computer Science (IJARCS), ISSN 0976 – 5697.
17. Tunitan, D. and Becker, K. (2014). Sentiment-based features for predicting election polls: a case study on the brazilian scenario. In Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on, volume 2, pages 126–133. IEEE.
18. <https://monkeylearn.com/sentiment-analysis/>; last accessed: 20 September 2018.
19. <http://blog.kaggle.com/2017/10/05/data-science-101-sentiment-analysis-in-r-tutorial/>; last accessed: 20 September 2018.