

POLITICAL SENTIMENT ANALYSIS ON DELHI USING MACHINE LEARNING

DHARMINDER YADAV ¹, AVINASH SHARMA, SHARIK AHMAD, AND UMESH CHANDRA

ABSTRACT. Social Networking sites are the best source of textual information which is used for emotion analysis and users view mining. Twitter tweets are used for sentiment analysis and analyses who will win the election of Delhi. Researchers collect the data from Twitter by using R API. We are using corpus based and Document based method for sentiment analysis, show the ten sentiments of tweets and classified these ten sentiments in to three sentiments such as Positive Negative and Neutral. RTextTools package is used to find the accuracy precision and recall. In RTextTools package total nine algorithms, but researchers used eight algorithm such as Scaled Linear Discriminant Analysis (SLDA), Support vector machine (SVM), Additive Logistic Regression (LOGIT-BOOST), Maximum Entropy (MAXENTROPY), Bootstrap Aggregating (BAG), Random forest (RF), Neural Network (NNET), Classification or regression tree (TREE) are used to for text classification. Navies Bayes theorem is used to find the Accuracy of algorithm and Precision, Recall for Positive Negative and Neutral tweets. This paper tried to show the opinion of Twitter user's tweets in English regarding Delhi Election by using Machine learning algorithm, Twitter, tidytext, tm package.

1. INTRODUCTION

Twitter text data best for sentiment analysis, because today most of the people spending their time on social networking sites and sharing their view on twitter

¹*corresponding author*

Key words and phrases. Twitter, sentiment Analysis, Natural language Processing, R, RText-Tool, Twitter API, Navies Bayes.

Facebook etc. Social Networking site users show their view regarding product service and political parties. This paper used the twitter text data of Delhi for political sentiment analysis. Scholar tries to find the mood of Delhi on the base twitter tweets. Twitter tweets contain 17 columns when collected from the twitter by using R API tool but we used only text columns for sentiment analysis. Political sentiment analysis use to analysis views of twitter user regarding political parties and know the belief, thinking and attitude of parties candidates or members. Researcher is classified the sentiment into positive, negative and neutral. They are used Natural Language Toolkit (NLTK) and Natural Language Processing (NLP) by using R to find the opinion of twitter user regarding Delhi election.

Delhi is capital of India and also called heart of India. Population of Delhi was in 2016 more than 26 million and second largest urban area of world [1]. Now boundary of Delhi extending to their neighboring cities such Gurugram, Noida, Faridabad, Panipat, Sonapat, and Ghaziabad is called National Capital Region (NCR). [2]. Most of the area of Haryana comes under region of NCR. Old name of Delhi was Indraprastha Indian epic Mahabharata [3]. Qutub minar and Red fort are listed in UNESCO heritage site. Delhi renames as 'New Delhi' in 1927 and it is also name as Lutyens' Delhi [4]. Delhi has its own Lieutenant governor, Legislative assembly, member of council and Chief Minister. In Delhi one division, 11 District, 33 subdivision, 59 census town, 5 Municipal corporation of Delhi (MCD) and 300 villages [5,6].

This paper finds the result of Delhi election 2020 and compares the election 2020 result with nine different algorithms such as Scaled linear discriminant analysis (SLDA), Support vector machine (SVM), Additive Logistic Regression (LOGITBOOST), Maximum Entropy (MAXENTROPY), Bootstrap Aggregating (BAG), Random forest (RF), Neural network (NNET), Classification or regression tree (TREE) and Navies Bayes Theorem and shows result [7] Precision, Recall, and Accuracy. This paper collected tweets from 1 January 2020 to 8 February 2020 that is available on twitter.

This paper organized as follows, Discuss introduction of Text mining, sentiment analysis, and Delhi in section 1. Literature review in section 2. Proposed methodology, pre-processing, and sentiment analysis of tweets in section 3. Next we discuss the result, classification of the polarity in graphical reorientation and

Machine learning algorithm in section 4. Section 5 contains the Conclusion of this Paper.

2. PRELIMINARIES

This section of paper explains the related study of Political sentiment analysis and related tools. Political sentiment analysis the twitter user opinion regarding political parties and try to predict who will be the winner in the Election. This paper tries to predict the outcome Delhi election on the bases of tweet collected from twitter in English language. Yadav et al. [8] start research on political sentiment and write paper on Indian election 2019, try to predict which party wins the election, understand the mood of Indian people. They also try to predict the election result of Haryana, opinion of Haryanvi twitter user regarding present government of Haryana in which they found sentiment was in favor of present government in Haryana result also come in favor of present government of Haryana.

Elghazaly et al. [9] start research to find the sentiment of people regarding presidential election in Egypt 2012. They study the text classification in Arabic language by using WEKA tool. They also used the Machine learning algorithm support vector machine (SVM) and Naïve bayesian theorem (NB) to find the accuracy, precision, and recall and compare their result with each classifier and find out NB classifier provide best result for Arabic text classification. Rushdi-Saleh et al. [10] have done sentiment analysis on movie review in Arabic Language. They used Arabic corpus to know the opinion of users regarding Movies. They collected 500 movies review in Arabic language from different site in which 250 review are positive and 250 reviews are negative. They utilized term recurrence (TF) and term recurrence opposite record recurrence (TF-IDF) for weighting reason. SVM and NB two machine learning classifier used to predict the accuracy, SVM provide better result when apply steaming and NB provide better result when don't apply steaming, but difference between SVM and NB are very less.

Malik et al. [11] analysis the twitter users view about Narendra Modi. They analyze the popularity of Narendra modi before election and after election. They found the popularity of Narendra modi increases after foreigner visit.

Godbole [12] proposed a News and blog sentiment analysis. They used WordNet to classify the news and blog into positive and negative polarity. They removed the unwanted term, text and found that machine learning method provide better result as comparison to simple method. Choy et al. [13] utilizes online conclusion to foresee the vote rate for every one of the applicants in the Singapore presidential appointment of 2011. Wang et al. [14] proposed a continuous assessment examination framework utilizing Naive Bayes model with unigram highlights for political tweets and gathered the estimation explanations utilizing Amazon Mechanical Turk which depended on the U.S. presidential appointment of 2012. Mishra et al. [15] uses Machine learning and NLP to know the opinion regarding digital India. Ahmad et al. [16] used Support Vector Machine algorithm of machine learning and categorized the text data polarity into Positive Negative and Neutral. They used grid search method to tune accuracy of SVM algorithm and other metrics such as Precision, Recall and F-measure.

Smeaton and Bermingham [17], uses tweets from Twitter, reviews from Blippr tested the hypothesis and analyze that sentiment analysis is easier in micro text as compared to longer documents. They used twitter tweets and movie reviews and find out that experimented with Twitter, blog entries and film surveys and inferred that it is simpler to recognize emotion.

Pak and Paroubek [18] proposed a model for sentiment analysis by utilizing twitter tweets and arrange the tweets as a Negative and Positive. By utilizing 'Twitter API' they made a twitter corpus by gathering tweets and consequently explaining those tweets utilizing emojis. The multinomial Naive Bayes slant classifier methodology was created utilizing that corpus. To anticipate the opinion Naive Bayes utilizes the component of POS-labels and N-gram.

Po-Wei Liang et al. [19] used Twitter API to gather tweets from twitter and sifted through the tweets feelings. Unigram Naive Bayes model was produced for extremity recognizable proof. They dispense with the undesirable highlights by utilizing the Mutual Information and Chi square component extraction technique. Be that as it may, the foreseeing the tweets as positive or negative didn't give better outcome by this technique.

N. Rochmawati and S C Wibawa [20] use twitter for sentiment analysis for Rohingya Muslim. They used R language to categorize the polarity of tweets as a negative, positive and neutral. Eman M.G.Younis [21] use twitter tweets for sentiment analysis on Tesco and Asda store of UK over Christmas period 2014.

They analysis the customer opinion on the base of review provided by users for product and services. Pennebaker et al. [22] developed software to extract the emotion from the tweet. LIWC is program that evaluates passionate, intellectual and basic segments of content examples utilizing a psychometrically approved interior word reference. Tumasjan et al. [23] focused on the German election in 2009 by using twitter tweet and predict election result and reasoned that the quantity of tweets/notices of a gathering is legitimately corresponding to the likelihood of winning the decisions. They gathered more than one hundred thousand tweets dating from August 13 to September 19, 2009 containing the names of the six gatherings. Caetano et al [24] proposed a political analysis on 2016 American Presidential Election by using twitter data. They collected 4.9 million tweets from August 2016 to November 2016. They find out the political homophily between twitter users regarding US Presidential Election and classify the tweet into three polarities Positive, Negative and Neutral. They analyses the homophily on three scenarios such as twitter follows, mention and Retweet connection, second analysis the multiplex connection third friendships with similar speeches.

3. METHODOLOGY AND FRAMEWORK

Political sentiment analysis is technique find the twitter users view regarding political party and leader or member of political party. This study finds view of twitter user regarding Delhi election by using text mining and opinion mining technique. Length of text message of tweets is 140 characters in which twitter users express their view regarding political parties [23]. Text mining extracts the hidden information from the text. Figure 1 Show the procedure used for sentiment analysis and text mining.

3.1. Data Pre-processing. Data collected from twitter by using twitter app and by using credential of twitter API. One lakh tweets collected from the twitter related to present government and chief minister (CM) of Delhi. Analyze the view of twitter user regarding cm of present government of Delhi. Twitter tweets contain 17 columns such as X, text, favorite, favoriteCount, replyToSN, created truncated, replyToSID, id, replyToUID, status Source, screenName, retweetCount, isRetweet, retweeted, longitude, and latitude. Figure 2 show the text

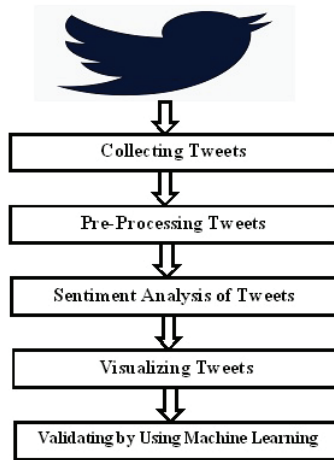


FIGURE 1. Overview of Methodology

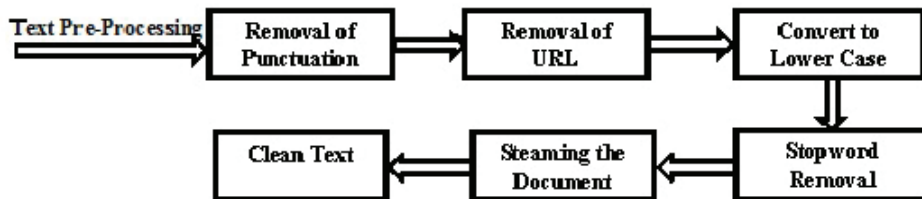


FIGURE 2. Text Pre-Processing of Text mining

pre-processing steps. After text pre-processing and removal of Duplicate tweets only 32301 tweets left for sentiment analysis.

4. RESULTS AND DISCUSSION

This part of paper, discuss the result of our experiments. Sentiment analysis is the process that categorized the tweet data into Positive, Negative, Neutral sentiment. Find the emotion of data by using tm package of R language. We categorized these tweets as Positive, Negative and Neutral on the base of polarity. If polarity of tweets greater than zero then tweets is positive. If polarity of tweet less than zero tweets is negative. If polarity of tweet is equal to zero then tweets are neutral. Figure 3 Show the 10 sentiment of Natural Language Processing (NLP). Highest numbers of tweets are positive, second highest numbers of tweets are trust and third highest number negative its means that people

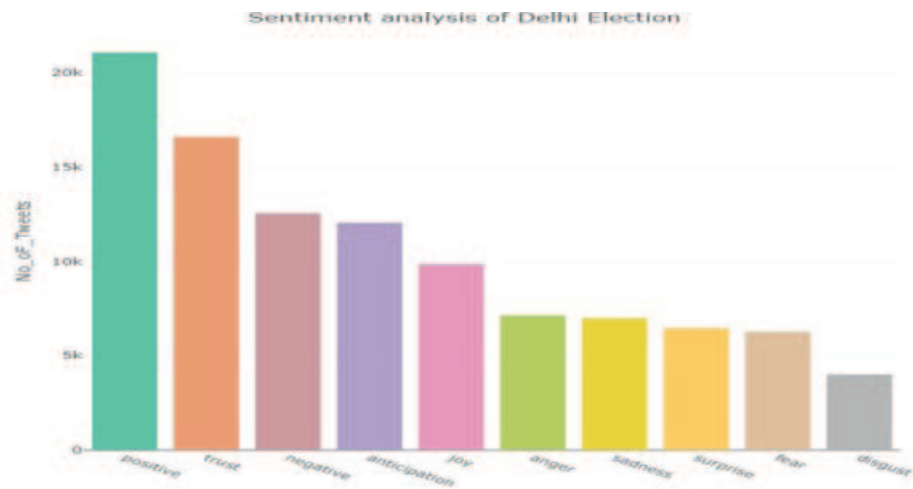


FIGURE 3. Tweets with 10 Sentiments

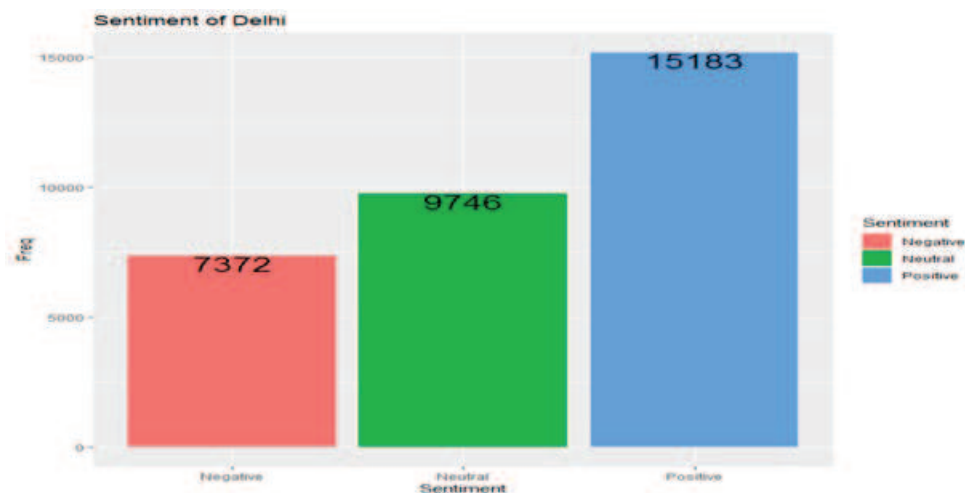


FIGURE 4. Positive, Negative and Neutral Number of Tweets

thinks positive from twitter user's point of view, other tweets are few in number as compare to positive and negative tweets. Figure 4 shows the three sentiments of tweets which classified from these 10 sentiments. Positive tweets are 15183, negative tweets are 7372 and neutral tweets are 9746, it shows that numbers of positive tweets are approximate more than 2 times of negative tweets. Figure 5 shows wordcloud with ten sentiments in different color. Congress, Vote, Good, Sir and Delhi word appear bold and bigger in size its means it that frequency

TABLE 1. Accuracy, Precision, Recall, and F-score of eight Algorithms

Algorithm	Precision	Recall	F1 score	Accuracy	Mean Accuracy After 4 Iteration	Mean Accuracy After 10 Iteration
SVM	0.87	0.72	0.78	0.93	0.9379	0.9360
SLDA	0.90	0.68	0.75	0.92	0.9233	0.9239
LOGITBOOST	0.84	0.72	0.77	0.94	0.9469	0.9446
BAGGING	0.85	0.71	0.76	0.93	0.9362	0.9363
RANDOM FOREST	0.87	0.75	0.80	0.93	0.9373	0.9371
TREE	0.87	0.60	0.68	0.93	0.9002	0.8987
NEURAL NETWORK	0.85	0.72	0.77	0.91	0.9250	0.9250
MAXENTROPY	0.82	0.74	0.77	0.90	0.9426	0.9446

TABLE 2. N-Ensemble Algorithm

N-Ensemble	N-Ensemble Coverage	N-Ensemble Recall
N>=1	1.00	0.92
N>=2	1.00	0.92
N>=3	1.00	0.92
N>=4	1.00	0.92
N>=5	0.99	0.93
N>=6	0.98	0.93
N>=7	0.96	0.93
N>=8	0.94	0.94
N>=9	0.89	0.96

5. CONCLUSION

This paper concludes twitter user view regarding Delhi election. By sentiment analysis found that most of the tweets are Positive regarding present government of Delhi, difference between positive and negative tweets are 7811, which is approximate 24 percent of all tweets and 22.15 percent of negative tweets. Positive tweets are 47 percent of total tweets its means that twitter users are in favorers of present government of Delhi and happy with present government

TABLE 3. Navies Bayes Theorem Precision, Recall and Accuracy

Accuracy	Class	Precision	Recall/ Sensitivity	F1 Score	Specificity	Balanced Accuracy
77.64	Negative	77.17	48.94	59.90	96.54	72.74
	Neutral	70.75	72.95	71.84	89.34	81.15
	Positive	80.79	90.03	85.16	74.27	82.16
	Total	76.23	70.64	72.30	86.72	78.68

work, result of Delhi on 08 February 2020 also comes in favorer of this experiment and sentiment analysis. NLP, NLTK and Bag of Word are used for sentiment analysis and text mining. Nine algorithms of RTextTools packages are help to find the accuracy of text data. Boosting algorithm gives best 94 percent, second best result given by SVM, Bagging, Random Forest, and Tree algorithm with 93 %, Tree algorithm result decreases with after 4 fold and 10 fold but result of SVM, Bagging, and Random Forest increases or stable with 4 and 10 fold. Navies Bayes Theorem Accuracy is 77.64% which is less than all eight algorithms. From the above experiment, found that twitter user think positive regarding present government of Delhi and result of Delhi also show that people vote to present government of Delhi. There is chance of improvement in this result due to data incompleteness and lack algorithms, next we try to implement other machine learning algorithm which provide better accuracy.

REFERENCES

- [1] THE WORLD'S CITIES IN 2016: *United Nations*. October 2016, 4.
- [2] DELHI (INDIA) *Union Territory, Major Agglomerations and Towns* , Population Statistics in Maps and Charts, 2017.
- [3] ECONOMIC SURVEY OF DELHI, 2005–2006, 2016: *Planning Department*, Government of National Capital Territory of Delhi., 1–7.
- [4] LUTYENS' DELHI IN RACE FOR UN HERITAGE STATUS : 11 June 2012, Hindustan Times.
- [5] M. S. A. RAO: *Urbanization and Social Change: A Study of a Rural Community on a Metropolitan Fringe*, Orient Longmans, 2012.
- [6] THE DELHI MUNICIPAL CORPORATION (AMENDMENT) *Act 2011 (Delhi Act 12 of 2011)*, 24 March 2017, Delhi.gov.in. Department of Law, Justice and Legislative Affairs, 2011.

- [7] P. J. TIMOTHY, L. COLLINGWOOD, A. E. BOYDSTUN, F. GROSSMAN, W. V. ATTEVELDT: *RTextTools: A Supervised Learning Package for Text Classification*, The R Journal, **5**(1) (2013), 34–45.
- [8] D. YADAV, H. MAHESHWARI, U. CHANDRA: *Political Sentiment Analysis on Indian Perspective Using Twitter Data*, Journal of Computational and Theoretical Nano science, **16** (2019), 4224–4231.
- [9] T. ELGHAZALY, A. MAHMOUD, H. A. HEFNY: *Political Sentiment Analysis Using Twitter Data*, ACM, ICC '16, March 22–23, Cambridge, United Kingdom, 2016,
- [10] M. RUSHDI-SALEH, M. MARTIN-VALDIVIA, L. URENA-LOPEZ, J. PEREA-ORTEGA: *OCA: Opinion corpus for Arabic*, Journal of the American Society for Information Science and Technology, **62**(10) (2011), 2045–2054.
- [11] A. MALIK, D. KAPOOR, A. P. SINGH: *Sentiment Analysis on Political Tweets*, International Symposium on Fusion of Science and Technology, New Delhi, India, 2016, 359–361.
- [12] N. GODBOLE, M. SRINIVASIAH, S. SKIENA: *Large-scale sentiment analysis for news and blogs*, In Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 2007.
- [13] M. CHOY, L. F. M. CHEONG, N. L. MA, P. S. KOO: *A sentiment Analysis of Singapore Presidential Election 2011 using Twitter Data with Census Correction*, Singapore Management University, Singapore, 2013.
- [14] H. WANG, D. CAN, A. KAZEMZADEH, F. BAR, S. NARAYANAN : *A System for Real-time Twitter Sentiment Analysis of 2012 US Presidential Election Cycle*, In Proceedings of the ACL 2012 System Demonstrations, 115–120.
- [15] P. MISHRA, R. RAJNISH, P. KUMAR: *Sentiment Analysis of Twitter Data: Case Study on Digital India*, International conference on information technology, 2016, 148–153.
- [16] M. AHMAD, S. AFTAB, M. S. BASHIR, N. HAMEED, I. ALI, Z. NAWAZ: *SVM Optimization for Sentiment Analysis* , International Journal of Advanced Computer Science and Application, **9**(4) (2018), 393–398.
- [17] A. F. SMEATON, A. BERMINGHAM: *Classifying sentiment in Microblogs: Is brevity an advantage?* Proceedings of the 19th ACM international conference on Information and knowledge management, 2010, 1833–1836.
- [18] A. PAK, P. PAROUBEK: *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*, In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, 1320–1326.
- [19] L. PO-WEI, D. BI-RU: *Opinion Mining on Social Media Data*, IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3-6. 2013, 91–96.
- [20] N. ROCHMAWATI, S. C. WIBAWA: *Opinion Analysis on Rohingya using Twitter Data*, IOP Conf. Series: Material Science and Engineering, **336** (2013), 23–35.
- [21] M. G. Y. EMAN: *Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study*, International Journal of Computer Applications, **112**(5) (2015), 44–48.

- [22] J. W. PENNEBAKER, C. K. CHUNG, M. IRELAND, A. GONZALES, R. J. BOOTH: *The development and psychometric properties of liwc*, 2007.
- [23] A. TUMASJAN, T. O. SPRENGER, P. G. SANDNER, I. M. WELPE: *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment* , ICWSM, 2010, 179–185.
- [24] J. A. CAETANO, H. S. LIMA, M. F. SANTOS, H. T. MARQUES-NETO: *Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election*, Journal of Internet Services and Applications, **2** (2018) 1–15.

DEPARTMENTS OF COMPUTER SCIENCE AND TECHNOLOGY
GLOCAL UNIVERSITY
SAHARANPUR, UP, INDIA
E-mail address: ydharminder@yahoo.com

DEPARTMENT OF CSE
MAHARISHI MARKANDESHWAR ENGINEERING COLLEGE
MAHARISHI MARKANDESHWAR (DEEMED TO BE UNIVERSITY)
MULLANA, AMBALA (HARYANA), INDIA
E-mail address: asharma@mmumullana.org

DEPARTMENT COMPUTER SCIENCE AND TECHNOLOGY
GLOCAL UNIVERSITY
SAHARANPUR, UP, INDIA
E-mail address: shariqgee@gmail.com

DEPARTMENT OF COMPUTER SCIENCE
BANDA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
BANDA, UP, INDIA
E-mail address: uck.iitr@gmail.com