International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 04, April-2019

Sentiment Analysis of Tweets for Indian Election Using Random Forest Classifier

Kiran Sangada[1], Jitendrakumar Dhobi[2]

[1]*P.G. Student, Department of Computer Engineering, Government Engineering College, Gandhinagar, Gujarat, India*
[2] *Associate Professor, Department of Computer Engineering, Government Engineering College, Gandhinagar, Gujarat, India*

Abstract— Now a day, a huge number of user opinions are posted on every social media by thousands of users for any events, topic, current news, etc. Twitter is one of the biggest social media where people can post their tweets, share, and discuss it. Sentiment analysis is a study of people's feeling, reviews represented in textual form. The election is one of the most popular topic on social media. User's feeling or review on a related election is improving the national democratic process of elections. The research work proposed machine learning-based classifier for classifying the Indian election-related tweets (Positive or Negative). We proposed hyperparameter tuned random forest classification which used the efficient feature to classify Indian election-related tweets. In this work, we compared the result with other different technique result in the comparison random forest technique gives the more accurate result. Our research aim is analysed user sentiments about the Indian election using random forest classifier.

Keywords— Sentiment Analysis, Hyperparameter, Classification, Machine Learning, Indian Election

I. INTRODUCTION

There are many platforms where millions of people can share their feeling, opinion, and review of different types of topics [1]. The platform like Facebook, Twitter where people can write their feelings, thought, etc. Twitter is a rich source of information for sentiment analysis [10].

330 million people are active on Twitter. Big prior to the election, investing by politicians in social media campaigns and debates and debates between their supporters and opponents only raise the claim that the views and ideas posted by users have an impact on election results. User reacts on twitter about Indian election party or candidate that can be used to understand how users react themselves. However, Sentiment analysis is the best way to analysis people reaction about Indian election party or party's candidate using Twitter data.

Sentiment Analysis is also considered as the use of text analysis to identify and analyze data based on the people feeling, review, etc. The main aim of sentiment analysis is to find the attitude and the polarity (Negative or Positive) with respect to any relevant topic or document [8]. Sentiment analysis can be observed at three levels, known as document level, sentence level, and entity level. At document level, whole document classifying into positive or negative sentiment. At sentence level, sentences analyzing and check whether the text contains positive or negative sentiment. And at entity level, a particular entity classifying.

There are two main techniques used in sentiment analysis to classify data (Positive or Negative) [7][9]. First is Knowledge-based approach that requires large dictionary database for identifying sentiments. And second is Machine learning techniques that used training set data to develop a classifier to classify sentiments. Machine learning approach is much simpler than Knowledge-based approach because there are not required a predefined database.

There are many classifier algorithms in machine learning approach [2]. Different machine learning-based classifiers have their own set of parameters, which are tuned before a model gets trained. These parameters are called as a hyperparameter. In this research, Random forest classifier used with tuned hyperparameters for classifying the Indian election-related tweets and compare accuracy result with accuracy result of other classifier algorithms.

This paper is organized in total VI sections as follows. Section II contains related work; Section III contains the work detail for which we have got an idea for improvement. Section IV contains our proposed methodology; Section V contains the result analysis and last Section VI has a conclusion.

II. RELATED WORK

There are many ways through which sentiment analysis can be carried out, but this paper only focuses on sentiment analysis using machine learning approach. The number of researchers used different classification model to classify sentiments but most of them are focused on Support Vector Machine, Naïve Bayes, Dictionary based and maximum entropy.

In [2], Brandon and Jing used a lexicon and Naïve Bayes Machine learning Algorithm to classify political tweets which are related to the 2016 US presidential election. They manually labelled a set of tweets for both Donald Trump and Hillary Clinton then compared their results with polling data. In result, Naïve Bayes Machine learning Algorithm is well performed then lexicon.

In [3], the authors used lexical based sentiment analyzer which classifies the Us presidential election tweets based on sentiment score. This classification is based on polarity and subjective measures that signify the positive, negative or natural attitude.

In [4], the authors compared the performance of two classifiers named Support Vector Machine and Multinomial Naïve Bayes using python package. Scikit learn and nltk that are python packages used for implementation. At the end of the result, SVM is batter performed with Scikit python package.

In [5], Ley and John have proposed a new method that considers neutral tweets related to a specific candidate's popularity which proved to increase prediction accuracy in the study of predicting the 2017 French election result.

III. EXISTING WORK

As specified in sector II, in a research paper [1] they have applied supervised approach and unsupervised approach on Indian election-related tweets which are in Hindi language. Naive Bayes (NB) and Support Vector Machine (SVM) classification algorithms are used as supervised approach. And Dictionary based algorithm is used as unsupervised approach. In dictionary-based algorithm, they have constructed SentiWordnet. SentiWordnet contained synonyms and antonym of the words along with the score of that particular word.

They collected a total of 23,998 Hindi tweets using Twitter Archiver. They collected tweets which used a hashtag of Indian political parties such as #BJP, #Congress, #NCP, #AAP. In pre-processing step, they removed website URLs, remove hashtags, twitter mentions (), stopwords, emoticons, and special characters and punctuations. They have used negation handling algorithm before using the classification algorithm. Negation handling algorithm handled negative words to help in finding the polarity of tweets. After finding the polarity of tweets, they have used three classifier algorithms such as Naive Bayes (NB), Support Vector Machine (SVM) and Dictionary based algorithm.

After applying classification algorithms, Naive Bayes classifier gave 62.1% accuracy and BJP 55% positive tweets had won the election in the year 2016. The accuracy of Support Vector Machine algorithm was 78.4% and BJP with 60% positive tweets had to win the elections in the year 2016. And dictionary-based algorithm gave 34% accuracy and Congress has 68% positive tweets.

After the analysis, SVM gave the higher accurate result and they predicted that the party that had a chance of winning the 2016 general election is BJP. The results are presented in Table I.

TABLE I
RESULT OF EXISTING SYSTEM

Algorithm	Accuracy	Precision	Recall
Dictionary Based	34%	-	-
Naive Bayes	62.1%	0.71	0.61
SVM	**78.4%**	**0.75**	**0.78**

IV. PROPOSED WORK

In our proposed work, we used hyperparameter tuned random forest classifier to classify Indian election-related tweets which are fetched from the Twitter API. We used different types of a feature as hyperparameter such as User's account details, hashtags, Mentions, Followers, retweets, Party. We tested the algorithm by changing these parameters manually and test again and gives higher accuracy than existing work. The flow work of the proposed work shown in fig 1.

Step of Proposed Method:

1. Fetch Tweets from Twitter API
2. Get Metadata from the User' account like Followers
3. Extract Feature From the tweets
4. Preprocessing tweets
5. Negation Handling like not, no good, etc.
6. Find Sentiment score
7. Select feature as hyperparameter
8. Apply Random Forest classifier

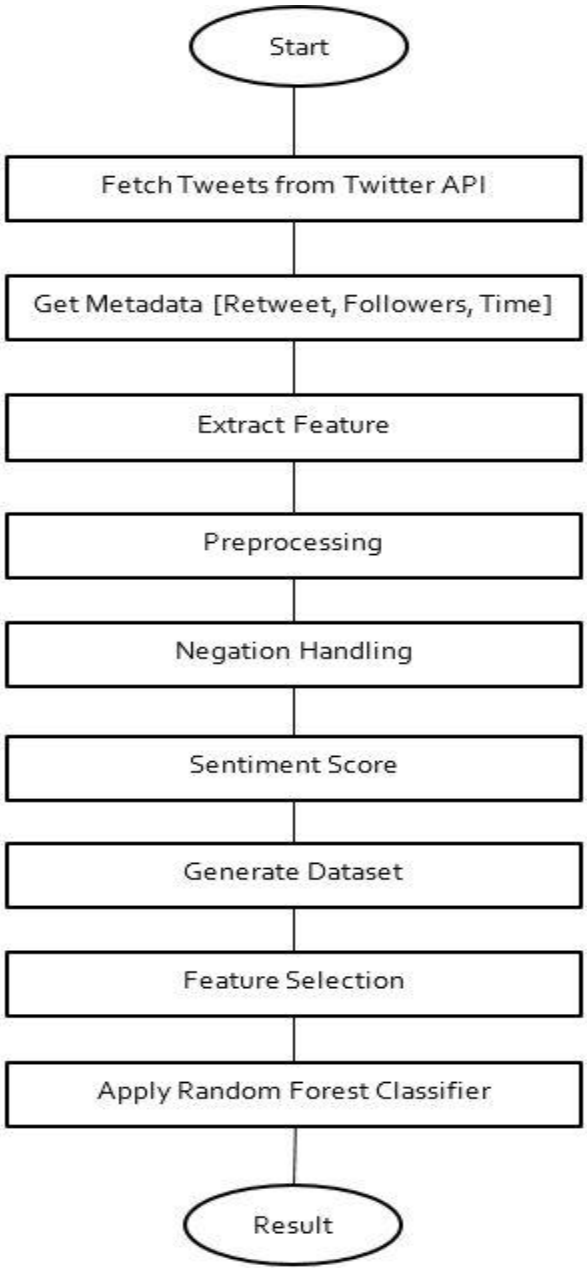


Fig. 1 Proposed Work Flow chart

V. RESULT ANALYSIS

We have applied hyperparameter tuned random forest classification to datasets as mentioned in Section IV. We have found some significant improvement in accuracy. We analyzed that the party that had a chance of winning the next election is BJP with 67% positive sentimental score. Table II contains proposed work accuracy compared with SVM algorithm. Fig 2 contain accuracy comparison with a different algorithm using a bar chart. As an analysis, our proposed method gives an overall 95.6% accuracy.

TABLE II
COMPARE PROPOSED ACCURACY WITH EXISTING ACCURACY

Algorithm	Accuracy
Support Vector Machine (SVM)	78.4%
Random Forest	**95.6%**

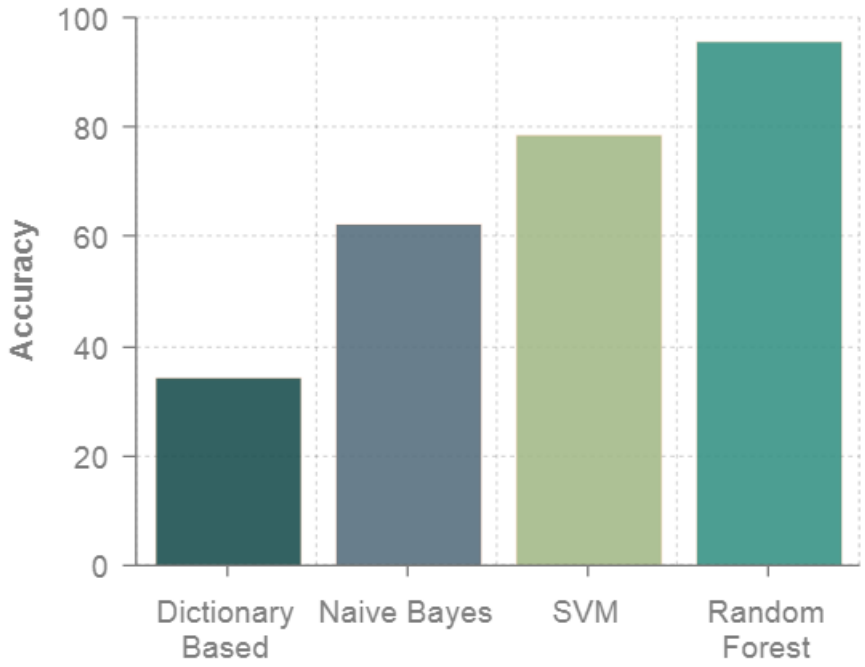


Fig. 2 Bar chart of Different Classifier Accuracy

VI. CONCLUSIONS AND FUTURE WORK

As shown as section v, the random forest classifier gives 95.6% accuracy that is higher than SVM. We sentiment that the party that had a chance of winning the next election is BJP. In future, there are many other social media like Facebook data used for classification. In future, we will used different classification algorithm to find more accurate result.

REFERENCES

[1] Parul Sharm, and Teng-Sheng Moh, "Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter," *in IEEE Access*, pp. 1966–1971, Dec. 2016.
doi: 10.1109/BigData.2016.7840818

[2] Neethu M S and Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques," *IEEE*, July 2013.
doi: 10.1109/ICCCNT.2013.6726818

[3] Brandon Joyce and Jing Deng, "Sentiment Analysis of Tweets for the 2016 US Presidential Election," *IEEE*, Feb. 2018.s
doi: 10.1109/URTC.2017.8284176

[4] Farha Nausheen and Sayyada Hajera, "Sentiment Analysis to predict election result using Python," *in IEEE Access*, pp. 1966–1971, June 2018.
doi:10.1109/ICISC.2018.8399007

[5] Jyoti Ramteke, Darshan Godhia, Samarth Shah and Aadil Shaikh, "Election Result Prediction Using Twitter sentiment Analysis, *"in IEEE Access*, Jan. 2017.

[6] Lei Wang and John Q Gan, "Prediction of the 2017 French Election Based on Twitter Data Analysis, *"in IEEE Access*, Nov. 2017.

[7] Bing Liu, Sentiment Analysis and Opinion Mining Morgan and Claypool Publishers, May 2012. .

[8] Y. Mejova, "Sentiment analysis: An overview," Comprehensive exam paper, available on "http://www. cs. uiowa. edu/˜ ymejova/publications/CompsYelenaMejova". pdf [2010-02-03], 2009.

[9] Walaa Medhat a, Ahmed Hassan b, Hoda Korashy b, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal science direct 2014.

[10] Malhar Anjaria, Ram Mahana Reddy Guddeti, "Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning," Sixth International Conference on Communication Systems and Networks (COMSNETS), IEEE, 2014.