# Twitter data analysis for Indian election

*Suraj Gupta*
*suraj885@ymail.com*
*IMS Engineering College,*
*Ghaziabad, Uttar Pradesh*

*Anish Anand*
*Anishanand119@gmail.com*
*IMS Engineering College,*
*Ghaziabad, Uttar Pradesh*

*Sudhanshu Singh*
*singhsudhanshusingh0525@gmail.com*
*IMS Engineering College,*
*Ghaziabad, Uttar Pradesh*

*Vaibhav Kapil*
*sharmavaibhav.3march@gmail.com*
*IMS Engineering College,*
*Ghaziabad, Uttar Pradesh*

*Mukesh Kumar Singh*
*mukesh.singh@imsec.ac.in*
*IMS Engineering College,*
*Ghaziabad, Uttar Pradesh*

## ABSTRACT

*Social media use is at an all-time historic high for India, so we considered one popular social media platform, Twitter, and tried to see if we could predict how a group of people felt about an issue by only analyzing posts from social media. For our research, we looked at tweets that focused on the 2019 Indian Prime Minister Election. Using these tweets, we tried to find a correlation between a person's tweets and whom he is going to vote. We wrote a program to collect tweets that mentioned one of the candidates, then developed a sentiment algorithm to see which candidate the tweet favored, or if it was neutral. After collecting the data from Twitter and comparing it to the results of the Electoral College, we found that Twitter sentiments corresponded with 73.8% of the actual outcome of the Electoral College. The overall sentiment of all tweets collected leaned more positively towards BJP than it did for Congress. Using the tweets that were collected, we also try to predict how different geographical locations affected a candidate's popularity.*

*Keywords*— *Bhartiya Janta Party (BJP), Indian National Congress (Congress), Sentiments, Twitter*

## 1. INTRODUCTION

In Indian politics, 2019 is an important year because it is an election year for the nation. The two presidential candidates were Narendra Modi for the BJP and Rahul Gandhi for the Congress. The Indian prime minister elections work using an Electoral College, meaning that each state has a number of electors who will vote for a candidate. Voters cast their vote for these electors, meaning that popular voter does not always decide the outcome. This makes predicting the election's outcome difficult, so leading up to Election Day, political scientists looked to polling data as well as computer analysis programs to try and see which candidate was poised to win.

Debates and conversations about the candidates and their policies most certainly happened verbally, but people also use social media to voice their opinions about the election. This is no surprise, especially due to the current usage of social media sites such as Facebook, with 1.79 billion monthly active users, and Twitter, with around 313 million monthly active users. In addition to these platforms, the rising popularity of sites such as Snapchat, with 150 million daily active users, and Instagram, with 500 million monthly active users, also contributed to social media conversations.

The main feature of Twitter is its use of hashtags, represented by a hash sign (#) preceding a word or phrase. These are used within tweets as a prefix by a user to identify different topics or keywords present within the tweet. This makes other users, could search other instances of that specific hashtag to see other tweets that match that topic.

Sentiment analysis of Twitter data is not a new phenomenon. Pang with his colleague Lee worked on creating algorithms to facilitate opinion mining and word analysis in their 2008 study. In 2009 a study was conducted by Go, Huang, and Bhayani to train a sentiment algorithm to detect a tweet's positivity about a certain subject by using emoticons. Pak and Paroubek continued upon this in 2010 by using the subjectivity and objectivity of words in conjunction with a tweet's structure to create a classifier that could use collected data to determine a tweet's sentiment. We decided to create our own sentiment algorithm that used a collection of words, each with its own sentiment value, to analyze tweets about the two candidates.

## 2. METHODOLOGY

Our methodology composed of collecting tweets from Twitter and performing a sentiment analysis on each tweet. Some of the basics methods which we followed can be listed as follows:

- Information on tools and vendors
- Fetching tweets
- Preprocessing the fetched tweets(Data Sampling )
- Data characterization
- Data Representation

Twitter is an important channel data for communication these days. Individuals share personal experience and thoughts, basically, they express their sentiments on this platform. Debates and conversations about the candidates and their policies most certainly happened verbally, but people also use social media to voice their opinions about the election.

Basically, Python 3.0.1 is used as the basic constructing tool for the coding areas and moreover, MongoDB is also used for structural data purposes. From 2016 - 2020 the number of twitter data ranges between 74-67% for the age group for less than 35 years and 24-33 % for the age group for more than 35 years.

Now the tweets can be fetched from the twitter by using Oath authentication in which we can simply fetch the tweets without login into twitter account. For this, Twitter will provide some keys and access token. Now the tweets are fetched by using the requests module of python. Fetched tweets are directly stored in MongoDB. As fetched tweets are in JSON format, so we store in MongoDB which support JSON format and from where the preprocessing of tweets can be done very easily. MongoDB provides the export feature.

Preprocessing of tweets is done so that we got all the tweets in the desired format. In preprocessing we eliminate those tweets which contain special symbols which c be processed.

After collecting the data, we analyze the tweets. First, we developed an algorithm to determine the nature of the tweet corresponded with. We have basically implemented the NLP and NAIVE BAYES algorithms.

Our search words were "mode", "Rahul_Gandhi", "BJP" and "Congress". We did not include "Gandhi" and "Narendra" as a search word because it is a common name and could result in too many results that did not relate to the election.
Finally, the result of the analysis is visualized in the form of different graphs such as bar graph.

## 3. PROPOSED SYSTEM

- Tweets Fetched
- Data Preprocessing
- Classification Algorithm
- Classified Tweets
- Sentiment in Graphical analysis

Our proposed system follows the above steps in sequential order.

## 4. SPECIFICATION

### 4.1 Hardware Requirements
- Laptop/Personal Computer with good internet connection up to 1-2 Mbps
- 8 GB RAM
- 500 GB to 1 TB Hard Disk

### 4.2 Software Requirements
- Windows 8/8.1/10
- MongoDB
- Twitter API Developer App

## 5. CONCLUSION

Here we have created sentiment analysis classifier for twitter data using labelled data set. Basically, negation detection is done for the sentiment analysis. Twitter is used globally by millions of users, and discussions on the social media platform can sometimes be used to see what the public's opinion is regarding a certain issue. After collecting tweets about the Indian prime minister election and analyzing them to determine how people viewed the two candidates, we saw that the sentiment according to Twitter was somewhat accurate. If sentiment analysis algorithms are improved and further developed, they could be used to predict election results in the future more accurately.

## 6. FUTURE WORK

- To increase the accuracy of the Support Vector Machine.
- Like English tweets, a classifier for Hindi tweets and another mother tongue could be developed.
- Applying semantic analysis to improve results.

## 7. REFERENCES

[1] N. Silver, "Election Update: The How-Full-Is-This-Glass Election," Last modified November 2, 2016, http://fivethirtyei ght.com/features/election-update-the-howfull-is-this-glass-election/

[2] "What happens if the US presidential election is close?" Reuters, last modified November 4, 2016, http://www.hindusta ntimes.com/world-news/factbox-whathappens-if-the-u-s-election-is-close/storyZpB0gvmqTVRkAkqtCYmZiK.html

[3] J. Uscinski, "The real presidential election is in December when the Electoral College votes," Last modified November 7, 2016, http://www.miamiherald.com/opinion/oped/article113171073.html

[4] S. Reich, "What will the US presidential election mean for Europe?" http://blogs.lse.ac.uk/europpblog/2016/11/01/trum p-Clinton-nato-Europe/

[5] "Number of monthly active Facebook users https://www.statista.com/statistics/264810/number-ofmonthly-active-faceboo k-users-worldwide/0

[6] "Twitter Usage / Company Facts" Twitter Inc., https://about.twitter.com/company

[7] "Number of daily active Snapchat users" https://www.statista.com/statistics/545967/snapchat-app-dau/

[8] "Number of monthly active Instagram users" https://www.statista.com/statistics/253577/number-of- monthly-active-Instagram-users/

[9] "Getting Started with Twitter," Twitter Inc., https://support.twitter.com/articles/215585

[10] "Posting a Tweet," Twitter Inc., https://support.twitter.com/articles/15367