

# Text-tone analysis of prominent people on social media

<sup>1</sup>Ayush Raina, <sup>2</sup>Tanmay Jain

<sup>1,2</sup>Student

<sup>1,2</sup> Department of Computer Science Engineering

<sup>1,2</sup>VIT University, Vellore, India

**Abstract:** Social networking sites are getting to be noticeably better through a great many clients. Twitter is a standout amongst the most well-known informal organizations that, clients can send short messages in particular tweet. Opinion analysis is utilized as a part of different fields to reach to a final opinion of a tweet. For the most part different internet business sites utilize this mentioned to enhance their customer support and value of organizations. In this paper, we have proposed an improved opinion analysis. The proposed technique outweighs the existing systems in metrics of accuracy. We have used a specific use case for the proposed opinion mining method which is the analysis of recent tweet data of prominent political leaders in the course of general elections. In addition to this, we propose an interactive method of collecting opinion and poll data, which in turn makes usage of all the data seemingly easy to use. Proposed system provides each and every aspect necessary for capturing a complete picture of the people's opinion and depict it in a way which will be helpful for people to understand where their leaders stand in the general elections.

**IndexTerms - Text Analysis, Opinion, Political views, Polarity, Influence.**

## I. INTRODUCTION

In the past couple of years, online informal community, for instance, Facebook, Twitter, etc. has ended up being one of the genuine ways for web customers to keep correspondences with their companions. As per reports, the amount of casual network customers has accomplished 2.1B until late 2018, and is assessed to be around 2.5B customers comprehensively, until the completion of 2019. Among these regions, Twitter has had the speediest improvement against other casual network goals. Twitter intends to allow individuals have association together through short message. Heartbreakingly, spammers use twitter as an instrument to send malevolent connections and messages to customers. The investigations exhibit that in excess of six percent of messages in twitter are spam. We can generally use paired characterization for spam acknowledgment that in the procedure, classes are stamped correspondingly. While at abnormality express, a predominant piece of one information gathering is a class and data out of class is an exception. The data stream is a great volume of changing data so data extraction ought to be resolved with one intersection upon information. So, it is sensible for using twitter condition that countless are created in the midst of multi day. In this article, we will consider spam as one quirk issue and recognizing comprehensive features of spammers and spam tweets and utilize a calculation of the information stream for spam tweets acknowledgment.

## II. RESEARCH METHODOLOGY

### EXISTING SYSTEM

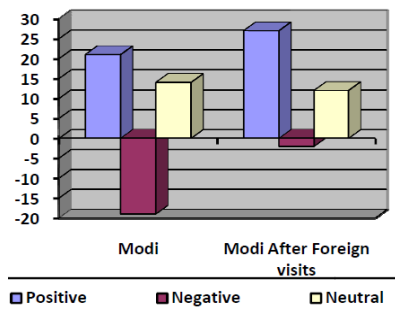
The current framework works just on the dataset which is obliged to a specific theme. The current frameworks likewise don't decide the proportion of effect the outcomes decided can have on the specific field contemplated and it doesn't permit recovery of information dependent on the question entered by the client for example it has obliged scope. In straightforward words, it takes a shot at static information as opposed to dynamic information. Unsupervised calculations like Vector Quantization, are utilized for information datasets and accordingly can create defective outcomes.

$$S_w = \frac{\sum_1^n S_{t(w \in t)}}{n}$$

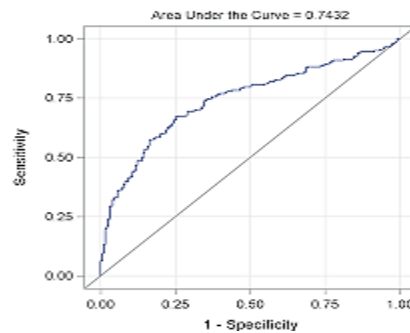
where

- N = no. of tweet that has the word
- St is sentimental score of that tweet that has the word "W".
- The sentiments of a tweet are determined on the basis of higher value of s+ or s-.

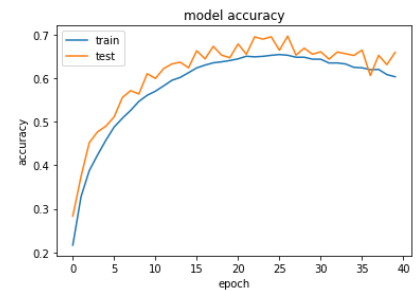
**Results of Existing System**



**Figure: Polarity score of tweets of Modi**



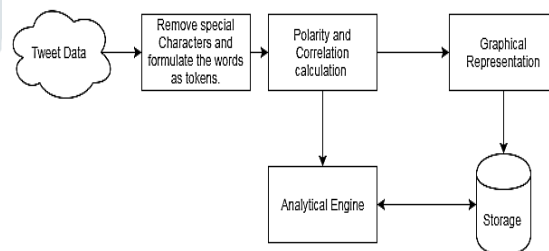
**Figure: Time Complexity Analysis**



**Figure: Accuracy Predicted on Actual data**

**III. PROPOSED SYSTEM**

We will use correlation on the tweet attributes to provide a comparative study to determine which political leader is popular among the mass. We will be using attributes such as the author-ID of the political leader, number of re-tweets for that particular leader, number of followers of the leader and so on. We can compute a correlation matrix for this scenario and using the matrix determine the sentiments of the people for that particular leader. In the proposed framework, we will recover tweets from twitter utilizing twitter API dependent on the question. The gathered tweets will be exposed to pre-processing. We will at that point apply the directed calculation on the put away information. The consequences of the calculations for example the slant will be spoken to in graphical way (pie diagrams/bar graphs). The proposed framework is more successful than the current one. This is on the grounds that we will probably know how the insights decided from the portrayal of the outcome can have an effect in a specific field.



**Figure 3.1: System Model**

**1. Creating Tokens**

First step is to create tokens for the tweets which can be used for analysis further in the model. We will create a tokens through removing any special characters, insignificant words such as the articles.

**2. Assigning Polarity and Formulating Correlation Matrix**

Every given token will have a certain polarity which can be accessed through the Database. These polarities will be used to assess if the people have a positive or negative influence of this particular search term. Correlation will be used determine various relations between the leader and its people and it will be used to improve the accuracy provided by the system by improving the positive-negative calculations.

**3. Analytical Engine**

It will be used to provide a proper analysis of our calculations. Will determine patterns through graphs and heat maps and according to people’s opinion will provide who is the best leader. It will store these analytical findings and compare it later to find if the image of that particular leader went down or went up.

**4. User level product**

Everything will be implemented in a concept interface for our polling platform, this is implemented in Adobe XD. It will provide a smooth functioning for our platform and we can also get a survey from people in that app about whom they think should win the elections.

## IV. IMPLEMENTATION

### Languages and Tools Used

Python 3

### Libraries

- Tweepy
- unidecode
- Textblob
- Matplotlib

### Technique & Algorithm used

Polarity Score  
Correlation

### Data Set

The data set consists of the set of the tweets posted on twitter obtained with the help of Tweepy.

### Keys

```
consumerKey = 'IKpPWsoDPIRclAqKS7MCjkDqH'
consumerSecret = 'ddVsHvD8UIG4oTszOAYMKZltpiWOB0ajf0ntDXG5QPOCVmaf5i'
accessToken = '1429632301-R4N1Ocha0QF7oWLGReAgKmiysqP2d9tXVfAOFZG'
accessTokenSecret = 'VpZTIQixRpwTqHH22hRypJLQkRWN0ezDtqXDn4RStN6bG'
```

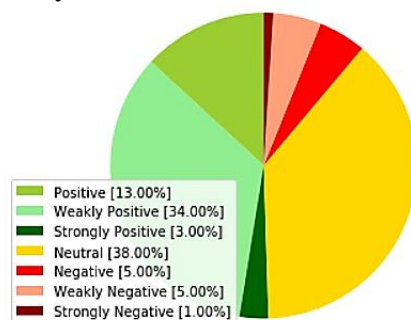
At first, we gather information from the twitter application utilizing catchphrases and hash names for the point to be analysed. By then convert the tweets into sentences and match each word with the game-plan of positive and negative words and consign positive to positive words and negative to negative words. By then add the score to every single sentence and understand the complete score. If the all-out score is more essential than zero it will be a positive tweet, negative tweet if the score is under zero and fair if the all-out score is zero. Directly, we processed the amount of positive, negative and objective tweets. By then we figure the measure of spam tweets. For this we utilize the going with features.

This paper takes a gander at the supposition examination of an unquestionable political figure by separating a dataset amassed from the scaled down scale blogging site page that is twitter. The dataset is gotten from twitter in the midst of the Prime Minister Narendra Modi's visit to United States of America. The endeavor work starts with the arrangement of a twitter application which supports us bring certain passage keys and tokens for the extraction and isolating of the tweets as required by the customer and putting it in a database for further mapping of these tweets for inclination examination of the proportional. The passage keys are delivered and the tweets are isolated and put in the database. The best approach to manage concentrate the tweets from the twitter database according to the customer's need is a traditional one as the customer has the control over what number of and which tweets are required by the customer and he/she can fill in according to the pinned for dimension of multifaceted nature.

### Backend System

It will be working on a python platform to provide seamless analytical tools to create statistical analysis of who will be winning the general elections.

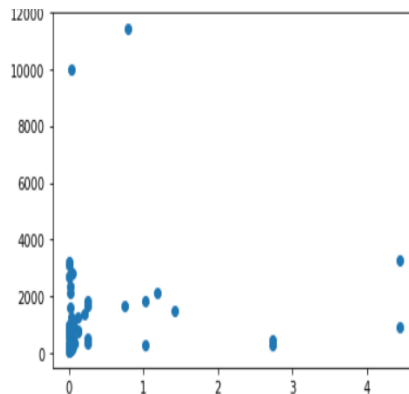
#### 1. Polarity and Correlation score analysis



**Figure 4.1: Showing analysis of a Leader**

This will depict the ratings of a particular leader in many ways like Positive, variations of Positive, Negative, variations of negative, Neutral. This graph will depict opinion of people for the particular leader.

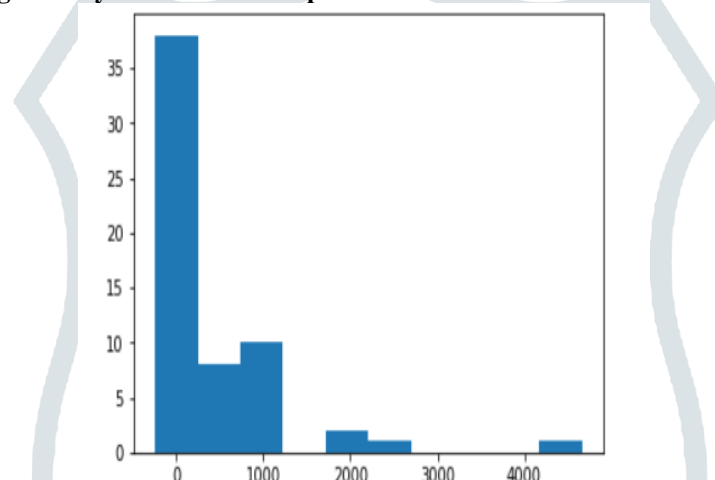
2. People correlation with the leader



**Figure 4.2: Correlation of authorID with followers**

This graph will provide an insight in the correlation of leader and their followers through a metric of finding how many retweets where done by the particular follower for their particular leader.

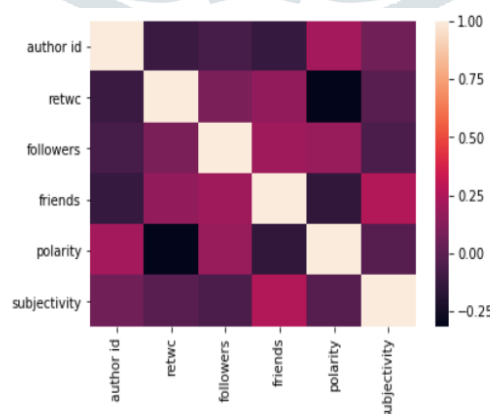
3. Tweets that are being liked by followers of that particular leader



**Figure 4.3: Histogram of likes for a tweet**

This graph depicts of many likes are the posts of that particular leader receiving. This could help us analyzes if the support of that particular leader is more or less.

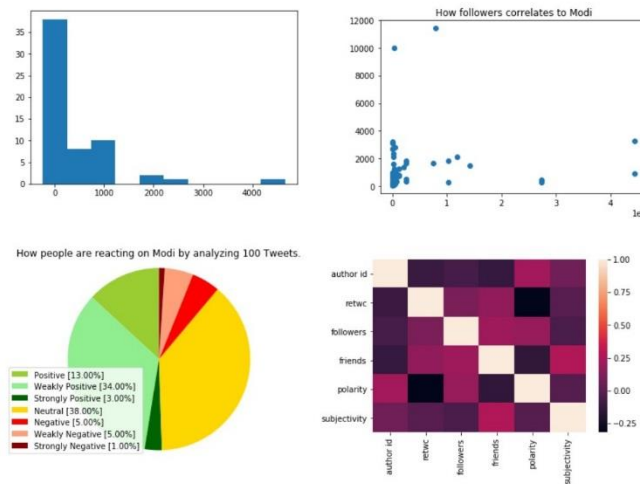
4. Heat Map



**Figure 4.4: Heat Map for Correlation Attributes**

This will represent important relations we need to look between author and followers to determine their popularity

**Example Illustrating Analysis of a Particular Leader – MODI**



**Figure 4.5: Analytical data for Hon'ble Prime minister - Narendra Modi**

**Frontend System**

Everything will be implemented in a concept interface for our polling platform, this is implemented in Adobe XD. This will provide a smooth functioning for our platform, where real users can provide correct and accurate data on which our text analysis can work upon. In addition to this, we can also get a survey from people in that app about whom they think should win the elections.

Here, we have showed the aforementioned application screenshots to our product:

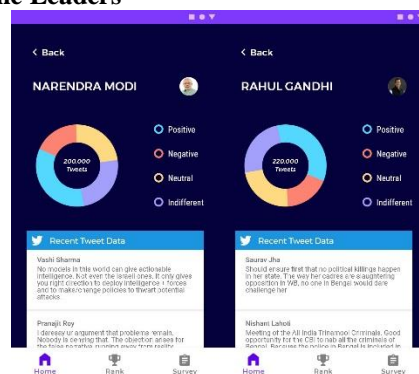
**1. Starting Pages**



**Figure 4.6: Depicts App Screenshots**

This provides an inside look of the user interface visualized for our sentiment analysis tool which shows the rankings of the political leader of the world. This will give our analytical engine the power to display the findings in an interactive manner for the people. People will understand the trend and can also answer survey questions to become informed in electoral decision making.

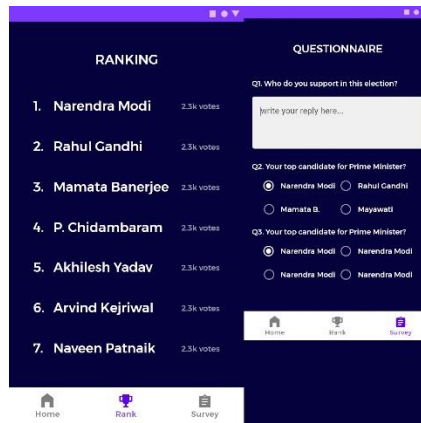
**2. Personalized information about the Leaders**



**Figure 4.7: Depicts App Screenshots**

These application screenshots depict a deep analysis on the particular leader. It provides us meaningful information why the ranking of a particular leader is high/low.

### 3. Ranking and Survey Page



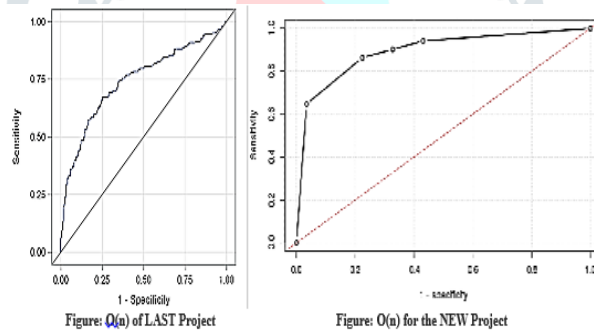
**Figure 4.8: Depicts App Screenshots**

These application screenshots depict the final rankings of leader according to people’s emotions. The second screenshot is survey screenshot which will be provided to people to get their personal opinion on our APP and use that data to further provide an accurate leader ranking. This concludes the implementation part of the project which depicts how our system analyses public data from twitter to provide leader ranking and graphical representation of the leaders.

## V. RESULTS AND DISCUSSION

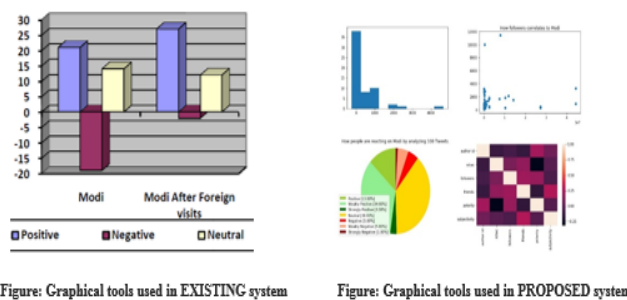
In this part we will discuss how our project was an improvement from the last implemented projects of the same ideas. We will provide a complete analytical comparison between the two and determine the best one which can be used in real life scenarios.

### 1. Time Complexity



1st represents the time complexity of the last project and 2nd one shows for the proposed one. We can clearly see that the graph shows that the specificity ratio of our project is greater than the previous project which depicts that time complexity of our project is lower than the existing ones. This is a plus point for us as our project is doing a high number of analytical operations on the tweet data and then in-turn using this to display meaningful information. Proposed system has a better time complexity.

### 2. Usage of Analytical data



These two figures represent the analytical data being used in two projects. We can clearly see the fact that we our using many more graphical tools to determine the popularity of the leaders. Proposed System has a better use of analytical tools.

### 3. Accuracy

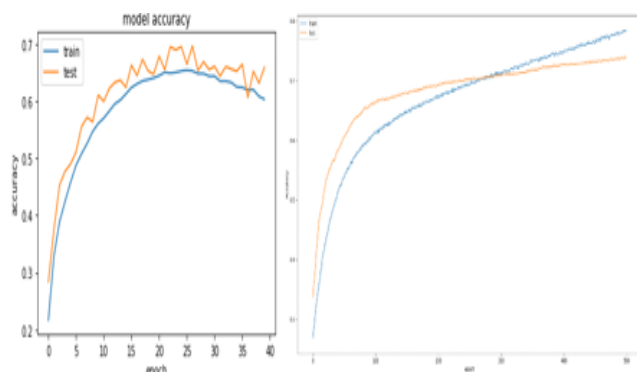


Figure: Accuracy of EXISTING system

Figure: Accuracy of PROPOSED system

These images depict the accuracy of our model. We can see the Accuracy of the project has improved many folds by adding correlation coefficient and using higher number of graphical tools to reach a decision. Proposed system has a better accuracy.

Sn	Existing System	Proposed System
1	Existing framework takes a put away informational collection on a specific subject into thought.	Proposed framework will give you the opportunity to pick the information of any subject.
2	It neglects to decide the effect the outcomes may or will have in the individual field.	Here, it gives you the effect the outcomes and measurements will have on the separate field.
3	Existing framework does not permit the recovery of information dependent on the inquiry entered by client.	Proposed framework permits recovery of information dependent on the inquiry entered by the client.
4	Existing framework does not give precise component determination.	Proposed framework will give exact element choice.
5	UI product services not provided	UI product services provided

Table 5.1: Comparison between the two systems

### VI. CONCLUSION

At first, we gather information from the twitter application utilizing catchphrases and hash names for the point to be analysed. By then convert the tweets into sentences and match each word with the game-plan of positive and negative words and consign positive to positive words and negative to negative words. By then add the score to every single sentence and understand the complete score. Directly, we processed the amount of positive, negative and objective tweets. By then we figure the measure of spam tweets. For this we utilize the going with features.

This paper takes a gander at the supposition examination of an unquestionable political figure by separating a dataset amassed from the scaled down scale blogging site page that is twitter. The dataset is gotten from twitter in the midst of the Prime Minister Narendra Modi's visit to United States of America. The endeavor work starts with the arrangement of a twitter application which supports us bring certain passage keys and tokens for the extraction and isolating of the tweets as required by the customer and putting it in a database for further mapping of these tweets for inclination examination of the proportional. The passage keys are delivered and the tweets are isolated and put in the database. The best approach to manage concentrate the tweets from the twitter database according to the customer's need is a traditional one as the customer has the control over what number of and which tweets are required by the customer and he/she can fill in according to the pined for dimension of multifaceted nature.

### Future Work

- **Including Machine Learning:** we will provide ML algorithm in our system to read the tweets and classify them using Naïve Bayes Algorithm or Natural Language Processing. This will provide a much better and accurate model and help achieve 100% accuracy in our model.
- **Global Implementation:** We our using the model to only help election process in India. We will be implementing this model for other countries and according to their country we can train different models which will suit that particular country.
- **Eliminating fake tweets:** We will be working on eliminating tweets which our being produced by bots or are displaying false information.
- **Providing a higher set of survey questions:** We will provide higher number of questions in the general survey to provide the candidate the person should select according to their choices in the survey.

## VII. ACKNOWLEDGMENT

The completion of this undertaking could not have been possible without the participation and assistance of so many people whose names may not all be enumerated. The contributions are sincerely appreciated and gratefully acknowledged. However, this group would like to express their deep appreciation and indebtedness particularly our guide and mentor, Dr. Lokesh Kumar R for his endless support, kind and understanding spirit during our work. To all the relatives, friends and others who in one way or another shared their support, either morally, financially and physically, thank you.

## REFERENCES

- [1] A. Goel and J. Gautam, "Real time sentiment analysis of tweets using Naive Bayes," *2nd International Conference on Next Generation Computing Technologies (NGCT)*, 2016.
- [2] V. Kalra and R. Agrawal, "Challenges of Text Analytics in Opinion Mining," *Advances in Data Mining and Database Management Extracting Knowledge From Opinion Mining*, pp. 268-282, 2019.
- [3] S. K. Paul and S. S. Rajput, "Applying Information Retrieval to Text Mining," *Practical Text Mining with Perl*, pp. 133-160, 2008.
- [4] B. Liu, "Sentiment Analysis and Subjectivity.," in *Handbook of natural language processing.*, Chapman and Hall, 2010.
- [5] M. Kaya, G. Fidan and I. H. Toroslu, "Sentiment Analysis of Turkish Political News," pp. 174-180, 2012.
- [6] D. Klein and C. D. Manning, "Accurate unlexical parsing," *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, vol. 1, 2003.
- [7] K. Mouthami, K. N. Devi and B. V., "Sentiment Analysis and Classification Based On Textual Reviews.," *IEEE International Conference on Information Communication & Embedded System*, 2013.

