SENTIMENT ANALYSIS OF SOCIAL MEDIA TREND FOR POLITICAL OPINION

Seeulee Upadhyaya[1] and Sourav Panigrahi[2]

Abstract—**Social Media has played a tremendous role inshaping public opinion in recent times. The magnanimous boom in the usage of social media with the emergence of multiuser platforms like Facebook, Twitter, Quora, etc; has provided necessary medium to share their opinions and views about almost anything around the world. The political scenario in India has not escaped the Internet Citizens or 'Netizens' who continue to share their political views openly on these platforms. However not all of the people who post these updates and read them have a very detailed view about Government policies and schemes. But the nature of the updates (read: the language, the tone, etc;) play a huge impact on the uninformed and regular netizens thereby shaping their opinion of a particular political party or candidate. This phenomenon, in turn, has a huge influence on voter preferences during the actual polling of votes. To project the effectiveness of Social Media on voter opinion a large dataset was cleaned, classified, and analyzed. The experimental results obtained have been used to decide upon the effectiveness of Social Media trends as an indicator of demographic opinion. The results can be also used to test the effectiveness of the online strategies employed by various political parties during the election season.**

Index Terms—**Social Media, Sentimental Analysis, Political, Facebook, Twitter, Quora**

I. INTRODUCTION

The magnificent growth in Social Media can be attributed to the development of user intensive platforms and various internet based applications that are designed to enable sharing of user content [1]. The advent of social networking sites and knowledge sharing platforms have enabled the masses to be able to share their opinion on various previously unheard of issues around the world. People these days are also able to gain insight on news around the world and that further shapes their opinion on these matters. This explosive growth can be seen on many platforms like Facebook, Twitter, Quora, etc.

Using technologies like Text Mining and Social Media Analysis, the huge volume on data on Social Media platforms, we can map user trends to their political intelligence and derive a relationship between how far a user influences or is influenced by the general opinion prevalent in the social media platforms.

II. RELATED WORK

Social Media Analysis has been a constantly growing trend and various research works have been published on the multitude of applications that may be sprung up from this analysis.

For example, Google researchers have demonstrated on their flu trends website how they could predict influenza epidemics two weeks ahead of the existing surveillance network through the analysis of query log data [2].

There has also been specific research on Political Analysis: Scharl, A., Weichselbraun [3] implements the US Election 2004 Web Monitor (webLyzard1), a specially designed Web portal that checked trends in political media coverage before and after the 2004 US Presidential Election.

The opinion mining prototype (OPTIMISM) developed by Carvalho, Sarmento have developed an opinion mining prototype system detects and classifies opinions about relevant political actors, regarding a particular topic of debate. Their system gathers opinion-rich texts from Portuguese social media, which are then classified according to their semantic orientation and intensity.

---

[1] *Dept. of Information and Communication Technology Manipal Institute of Technology Manipal University Manipal, India*

[2] *Dept. of Information and Communication Technology Manipal Institute of Technology Manipal University Manipal, India*

III. DATA COLLECTION

The collection of data was done through a survey on an open Google Form. The attributes selected were:
1. Name
2. Educational Qualification
3. 3 Policies/Schemes introduced by the UPA Govt
4. 3 Policies/Schemes introduced by the current NDA Govt
5. 3 Words to describe Mr.Rahul Gandhi
6. 3 Words to describe Mr.Narendra Modi
7. How often do you use Facebook?
8. How often do you use Twitter?
9. How often do you use Quora?

The data attributes aimed to analyze the political intelligence of the participant and the trend in their social media use. The results obtained from these attributes were also used in the classification of the records into Training Set and Testing Set.

The use of one word descriptions was useful since it helped us get a crisp view of the opinion and not check for any satirical or sarcastic tone in the input. It also helped in setting up Polynomial Scoring base to allocate rank scores to the records in terms of usability in our model.

IV. DATA CLEANING

In spite of using a 'pseudo-structured' data set a significant amount of data cleansing had to be performed to ensure that the classification can be performed in a sharper manner.

The process involved in cleaning included segregating the attributes '3 Words to describe Mr. Rahul Gandhi','3 Words to describe Mr. Narendra Modi','3 Policies introduced by the previous UPA Govt.' and '3 Policies introduced by the current NDA Govt.'.

The process also involved assigning NA to the attribute fields that remain blank e.g. if a participant had mentioned only 2 policies for the previous UPA Govt. the third column for this particular record would remain blank. It had to be converted to NA to facilitate scoring purposes later.

V.   DATA CLASSIFICATION

A.   *Calculation of Political Intelligence*

A new attribute 'political_intelligence',which helps to classify the participants was defined in the data set. The political intelligence for a candidate was calculated using the no. of valid policies submitted by the participant for each party collaboration (viz. the UPA and the NDA).

The index for scoring was 5 for 3 valid policies, 3 for 2 valid policies, 1 for 1 valid policy. The calculation was done for both the party collaborations and the sum gave us the political_intelligence.
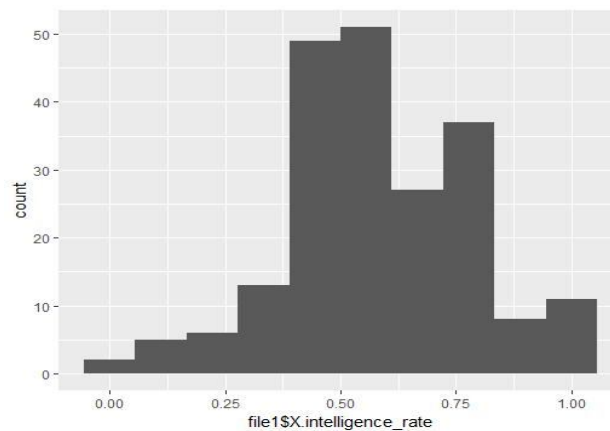

B. Calculation of Social Media Index

A new attribute 'social_media_index' was also defined in the dataset. This attribute helps us to classify the participants based on their social media presence and the regularity of use of social media platforms.
The various classifications were

- Everyday (Score 4)
- Once in 3 days (Score 3)
- Once in a week (Score 2)
- Once in a month (Score 1)
- Never (Score 0)

The sum of these scores for the three platforms for each participant gave us the 'social_media_index' for each participant.

'
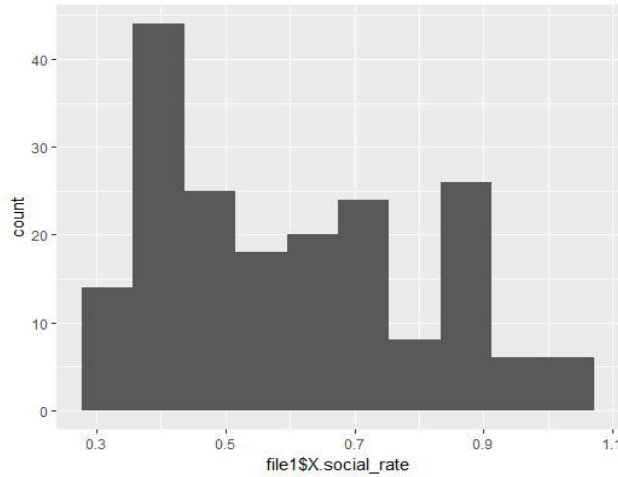
C. Normalization

The political_intelligence and social_media_index were then normalized using the maximum political_intelligence and maximum social_media_index.
For example,
Max(political_intelligence)=10
Max(social_media_index)=12
pact of each word on the psyche of each participant. This dataset was also cleaned further to include the small variations in the data elements e.g. 'irresponsible' in some responses became 'irreaponsible'. These minor changes were included in the scoring list to enable the system to accommodate for unintentional grammatical errors.
The response of each participant in the set was then checked against this scoring data set to calculate the 'Modi_Score' and 'RahulGandhi_Score' for each candidate.
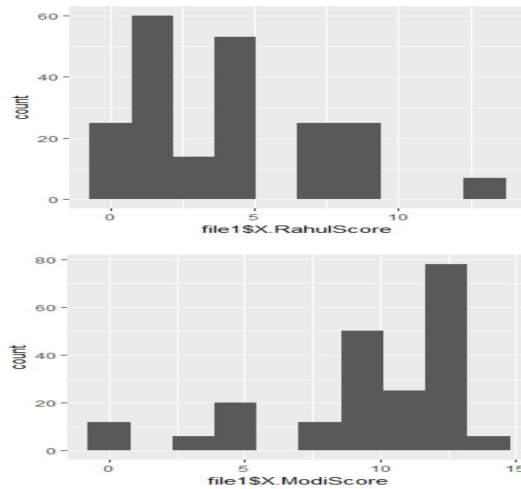

These scores were then used to further classify the candidates into 'Decided' and 'Undecided' voters. The

Decided voters formed the Training Set and the Undecided Voters formed the Testing Set.

VI. DATA MODELING

The Training Set data was modelled using a Decision Tree to obtain the parameters for finding the selection criterion for splitting the date in the Decision Tree.
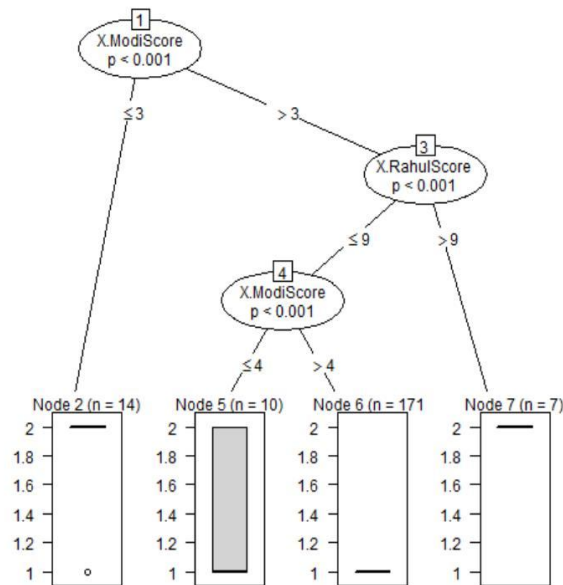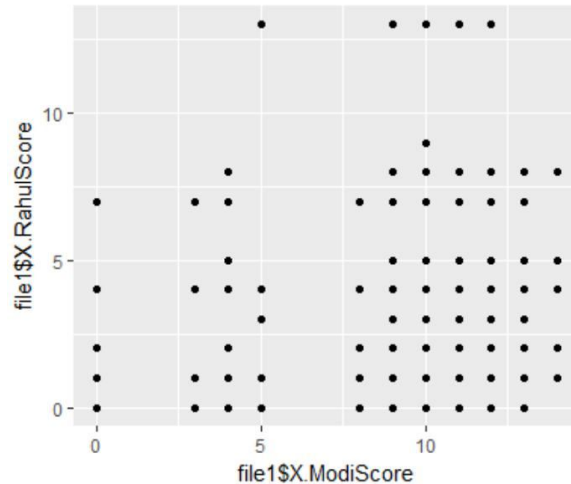
The modelling was done using RapidMiner tools and R Studio to obtain the classification points. The criteria used by political_intelligence,social_media_index,Modi_Score and RahulGandhi_Score. The results obtained here can be used to find the criteria that may be used to check the result for the undecided voters in the testing data set.

VII. CONCLUSION

Using the system developed we have calculated the number and distribution of supporters of Mr. Narendra Modi to be significantly more than the number and distribution of the supporters of Mr. Rahul Gandhi which is in tune with the results of the General Elections 2014.

REFERENCES

[1] K. Andreas, H. Michael, Users of the world, unite! The challenges and opportunities of social media, Business Horizons, Vol. 53, Issue 1 (page 61), 2010.

[2] Ginsberg J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S.,Brilliant, L. Detecting influenza epidemics using search engine query data, Nature 19 Nov 2008, 10.1038/nature07634, Letters to Editor, 2008

[3] Sarmento, L. and Oliveira, E. C. The Design of OPTIMISM, an Opinion Mining System for Portuguese Politics, In Portuguese Conference on Artificial Inteligence EPIA 2009,http://www.scientificcommons.org/51946156

[4] *Election Commission of India*

[5] S. Asur, B. A. Huberman, Predicting the Future With Social Media,International Conference on Web Intelligence and Intelligent Agent Technology, 2010.