International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

Sentiment Analysis of Twitter Data to Analyze the Acceptance of Aadhaar

Aarushi Thakral[1], Vasantha W. B.[2]

[1]SCOPE, VIT University, Vellore, Tamil Nadu, India

[2]Professor, SCOPE, VIT University, Vellore, Tamil Nadu, India

Abstract: ***India is a developing nation and investing 1.4 billion US Dollars is bound to have a huge impact on the economic health of the country, more so, if it has been invested into a single project, namely the Aadhaar project being implemented by UIDAI. However, it has been touted that this investment will pay itself off in terms of its applicability in various services, governmental as well as non-governmental. This project aims to analyze the general public perception of Aadhaar across the nation. Moreover, the twitter dataset being processed has been divided into two categories namely: developed and under-developed, on the basis of Gross Domestic Product (GDP) of the States and Union Territories of India. Equal number of tweets have been studied from both the categories of States and Union Territories to be able to place both the categories on a level field. The basic phenomenon of the needs of developing states differing from those of underdeveloped states has been exploited, as such sentiments would be expressed in the tweets from those locations. These tweets are assumed to reflect the opinions of the people living in that particular region. After close analysis, the findings of this project have been represented using various visualization techniques.***

Keywords: UIDAI, Aadhaar, GDP, Sentiment Analysis

1. Introduction

In 2009 the Government of India introduced a biometric identification scheme AADHAAR, for all the citizens of the country. It is a 12 digit unique identification number which contains the information of the residents based on their demographic and biometric data.

The information is collected by the Unique Identification Authority of India (UIDAI) established by GOI under the jurisdiction of Ministry of Electronics and Information technology following the provisions of the Aadhaar (Targeted Delivery of Financial and other Subsidies, benefits and services)Act, 2016 [2] [7]. Before the establishment of this statutory authority, UIDAI was functioning as an attached office of the Planning Commission which is now NITI Aayog. The budget of the project was approximately Rs. 8,793.9cr ($1.4 Billion) as of 30 Nov. 2017.[6]By the end of the year 2017, 1.19 billion citizens had an Aadhaar card which is almost 99% of the 18 and above population of India. The GOI made it mandatory for the citizens to link their bank accounts as well as mobile numbers, also public subsidy benefits and unemployment benefits can be availed only if you have an Aadhaar card; The Supreme Court in 2017, lifted its earlier ruling made in 2013 against making it mandatory, instead the ruling went towards making it mandatory for income tax, bank accounts etc. but voluntary for welfare schemes.

The court hearing held on 17th January, 2018 the announced that the last day to link Aadhaar to the banks and numbers will be 31st March, 2018, thus making it mandatory for majority of the government schemes.[8]

Some might compare it to the Social Security Number used in the U.S but it has more uses and less safeguards when compared. This in return has raised issues with people, raisings concerns about their privacy, making them skeptical about linking their bank details with it..Many residents in our country do not have any verifying documents causing them to not be able to access basic identity proofs like a ration card or bank passbook, Aadhaar solves this by being one sole identity card that can be used for almost any kind of verification and getting the necessary subsidies to the ones in need.[10]Also solving the downside of ration cards, there were people who used to make fake identities to get subsidies more than once Aadhaar solved this problem as subsidies are given using fingerprints rather than a card. Aadhaar in spite being helpful causes security concerns which a lot of citizens are not willing to afford. Every second day we hear a new piece of information regarding the disadvantages, every day we hear someone who disagrees with the methodology of the GOI regarding Aadhaar or someone who appreciates the help they are getting cause of it; this can be due to the fact that they belong to different classes, earn different amount per day or due to the fact that their state isn't that developed and hence they are not getting proper benefits. [1]

2. Project Process

2.1 Data Mining

For the mining of the data a registered twitter account was used. On the Twitter Developer site there is an easy ways to create a web application which helps us to download the data needed using a third party application, in this case R studio. [9]

After creating the API, we were provided with the following – 'Consumer Key (API key), Consumer secret (API Secret), Access Token, Access Token Secret.'

The *'twitteR'* package was needed to as it provided an interface for the Twitter API. [5]After all this the *setup_twitter_oauth (consumer_key, consumer_secret, access_token, access_secret)*, was used this was used to wrap the OAuth authentication handshake function from

their *httr*package to twitteR session. This allows R studio to search data with the help of *search_twitter()* package.

The data was taken from 29 states and 3 union territories. A total of 50,000 tweets were extracted on the basis of their longitude and latitude. The tweets from each state were different in number due to their population and area and hence they have been accordingly extracted.

The tweet extracted could be from the past 7 days only according to the new Twitter guidelines and hence were taken from the period of 14th January, 2018 to 20th January, 2018. There were certain parameters passed with the *search_twitter()*[4] package. These were:-

- Search Phase- which was set to "aadhaar" for all the tweets.
- Number of tweets- the number was different depending upon the area and population e.g- Pudducherry had 300 tweets whereas Delhi had 4000 for the same period of time.
- Language- was set to English using the keyword "en".
- Geocode- contained the latitude, longitude and the radius in miles.
- Date- to and from i.e. "14-01-2018 to "20-01-2018".

These tweets were in a form of tweet list and hence for easier access and faster use of R studio were converted to a data frame by using – *"twListToDF"*. This is included in the twitteR package.

2.2 Pre-processing

This step mainly deals with cleaning of the data for easier use in the analysis methodology. The tweet lists that are extracted have in total 19 columns including but we need only 1 the one containing the text. The row containing the tweet also had to be cleaned as the texts contains emoticons as well as links. The emoticons were not used as means of analysis as an emoticon can mean various things as a crying emoticon can mean crying due to laughter or due to sadness, even the smiling emoticon can mean sarcasm or happiness, the emoticons once transferred in R get converted to Latin, they were then converted to ASCII and then substituted with " " , i.e. a blank space. This was done using the *sapply()* function.

The hyperlinks and tags were removed using gsub() .We used the regular expression of Hyperlinks – " (f|ht)tp(s?)://(.*)[.][a-z]" and substituted them with "" i.e. blank space. [3]

The cleaned text columns are stored in separate data frames.

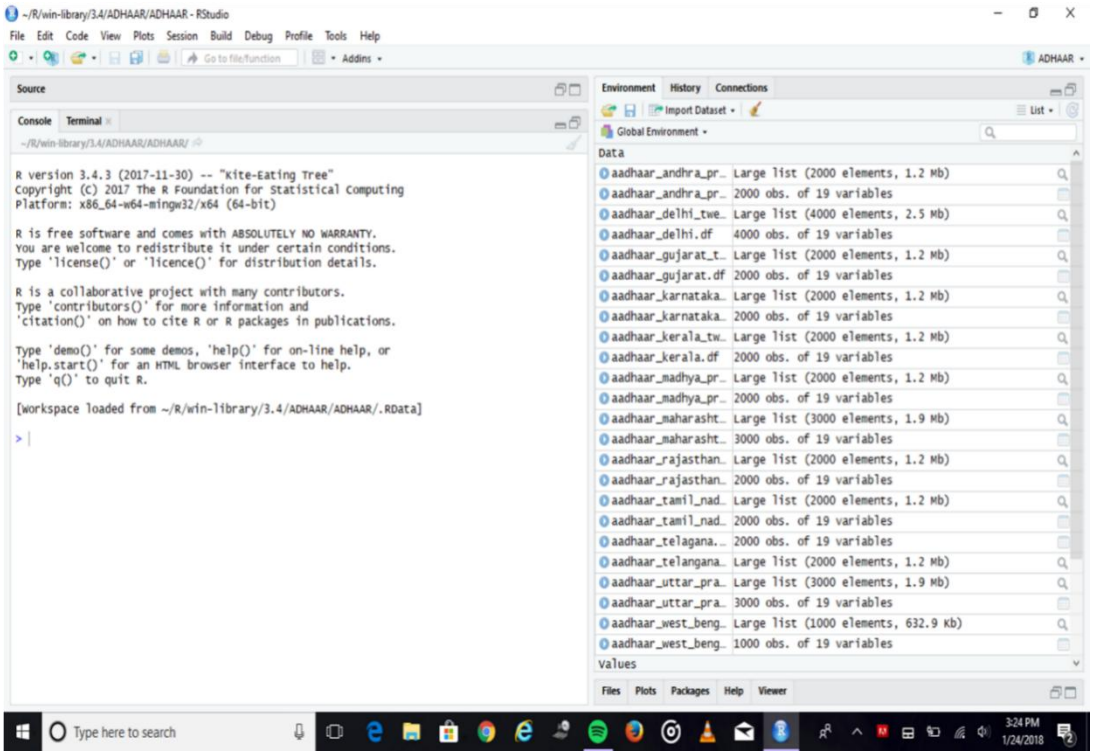

Figure 1: R studio Environment with tweet list state wise

2.3 The third step is Processing

1) In this step we first downloaded a dataset of positive and negative words. The datasets were then loaded to R studio for further use. The positive words dataset contains 2007 words and the negative words dataset 4873 contains words
2) The next step is to count the number of positive and negative words in each tweet of each state.
3) The positive words first were cross referenced by the tweet and then the count was put into the matrix and then done for all the tweets of that state the same step was repeated for the negative words. The date was then converted to a data frame for easier use and less usage of space.

4) The sum of the positive and negative words in the tweets of a state and presented in the form of a matrix the same matrix also contains the total score –

Total Score = Positive score – Negative Score

If the Total score of a state is negative the state is considered Against AADHAAR where as if the total score is positive it is considered for AADHAAR the following information is also put up in a matrix form using R language.

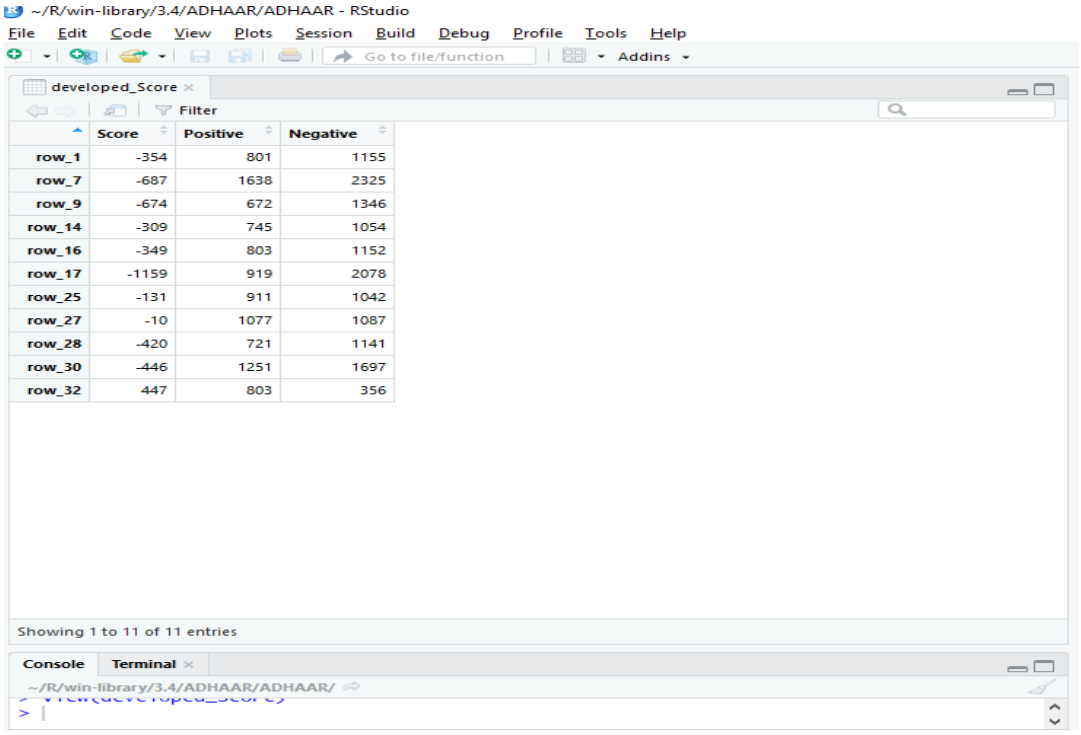

Figure 2: Row wise Developed state scores.

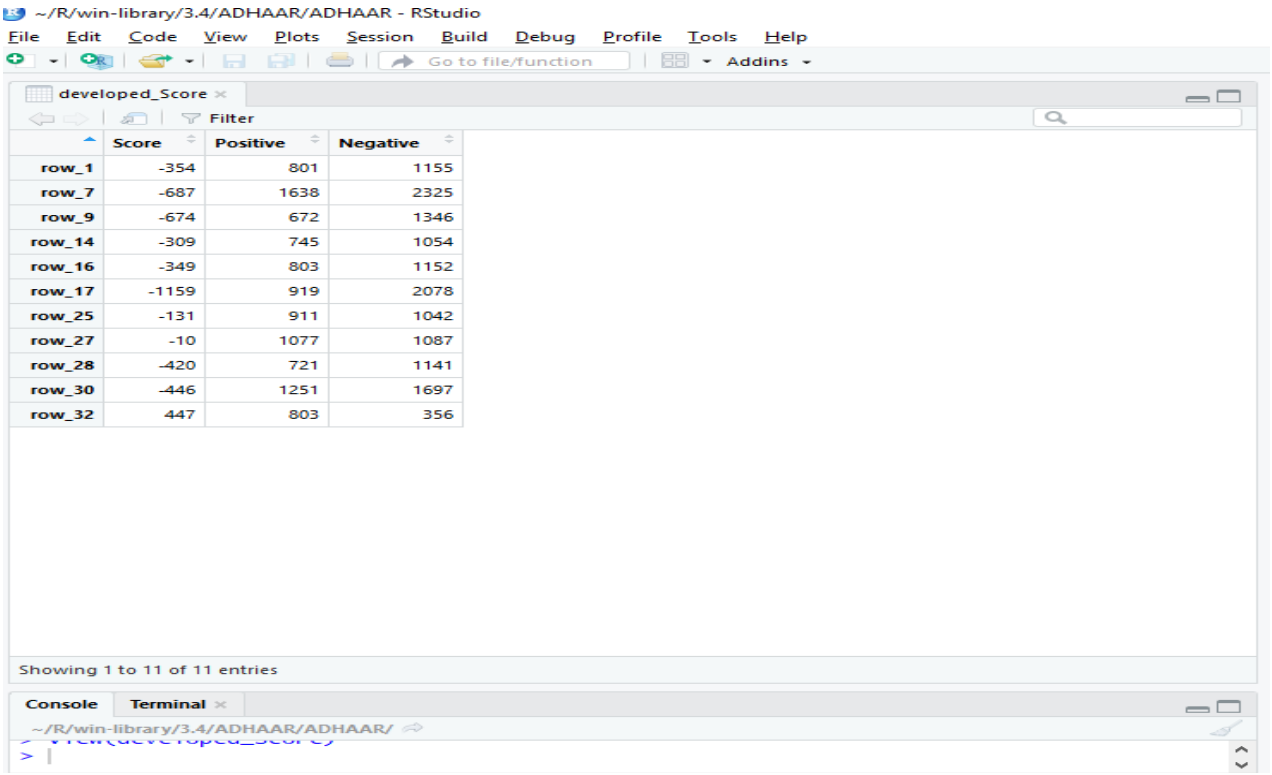

Figure 3: Row wise Underdeveloped state score

	State_column	V2
1	Andhra Pradesh	Against AADHAAR
2	Arunachal Pradesh	For AADHAAR
3	Assam	Against AADHAAR
4	Bihar	For AADHAAR
5	Chandigarh	For AADHAAR
6	Chattisgarh	Against AADHAAR
7	Delhi	Against AADHAAR
8	Goa	Against AADHAAR
9	Gujarat	Against AADHAAR
10	Haryana	For AADHAAR
11	Himachal Pradesh	For AADHAAR
12	Jammu Kashmir	Against AADHAAR
13	Jharkhand	For AADHAAR
14	Karnataka	Against AADHAAR
15	Kerala	Against AADHAAR
16	Madhya Pradesh	Against AADHAAR
17	Maharashtra	Against AADHAAR
18	Manipur	Against AADHAAR
19	Meghalya	For AADHAAR
20	Mizoram	For AADHAAR
21	Nagaland	Against AADHAAR
22	Odissa	Against AADHAAR
23	Puducherry	Against AADHAAR
24	Punjab	For AADHAAR
25	Rajasthan	Against AADHAAR
26	Sikkim	For AADHAAR
27	Tamil Nadu	Against AADHAAR
28	Telangana	Against AADHAAR
29	Tripura	For AADHAAR
30	Uttar Pradesh	Against AADHAAR
31	Uttarakhand	For AADHAAR
32	West Bengal	For AADHAAR

Figure 4: List of states showcasing their views towards Aadhaar.

2.4 The fourth step is Visualization

The visualization is done using various methods-

1) Plotting states with positive score with blue and negative score with red on the map.

Figure 5: Plotting states with positive score with green and negative score with red on the map

This image shows that only 13 States/UTs have positive score, colored blue whereas the other 19 have negative scores shown by Red color covering most of the map.

2) Plotting the Total scores on a bar plot.

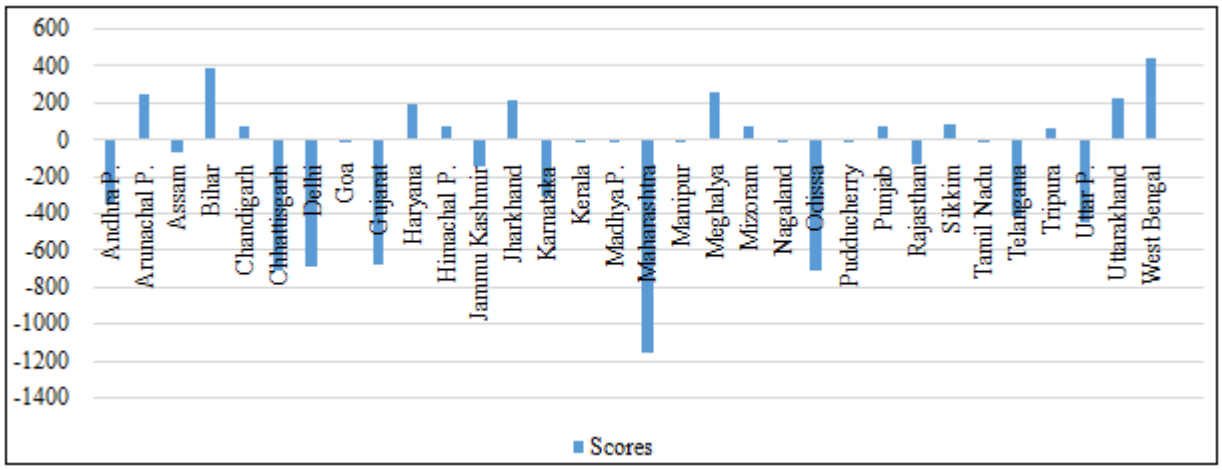

Figure 6: Plotting the Total scores on a bar plot.

The barplot displays the Total scores of each state i.e. the positive minus the negative score and UT plotted, according to this it is obvious the scores are very varied hence demonstrating ranging sentiments of the people.

The scores varies from 600 as that of West Bengal to -1150 as that of Maharashtra.

3) Plotting negative as well as the positive sore of tweets of all the states.

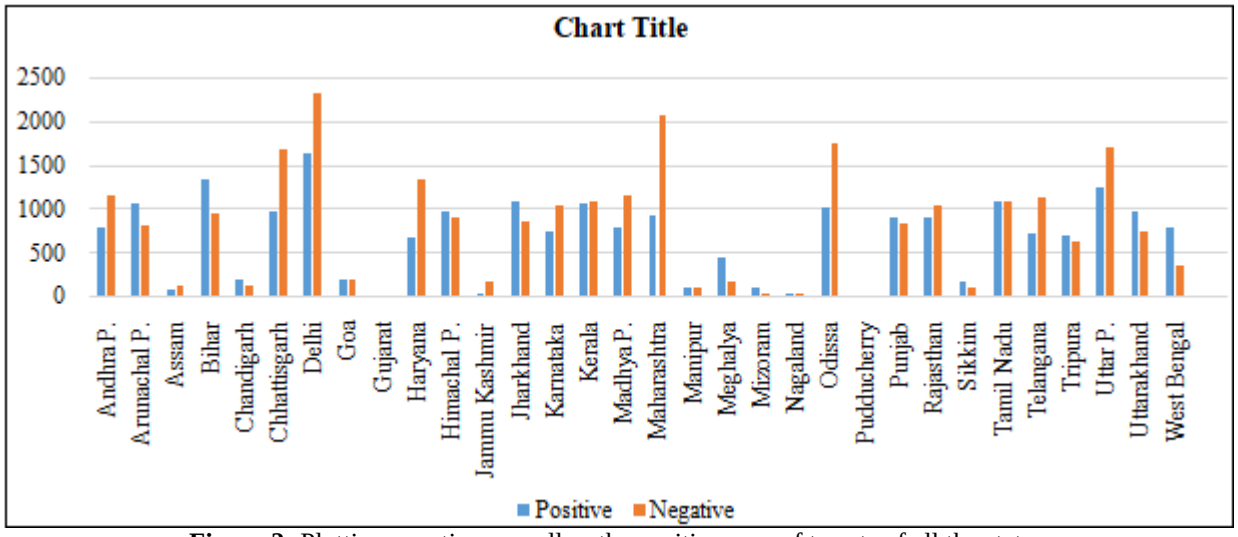

Figure 2: Plotting negative as well as the positive sore of tweets of all the states

This barplot illustrates the positive as well as the negative scores of the states the range here imparts that some states gave out less amounts of tweet as compared to others, hence the difference in the scores. In most of the states the Red being the negative is higher than the blue bar making the results obvious.

4) Plotting developed states with green and underdeveloped with purple on the map of India.

Figure 8: Plotting developed states with black and underdeveloped with gray on the map of India

This map illustrates that even though the developed states are only 11 they cover most of the area of India and also

showcasing how much their views matter and affect the views of the country.

3. Conclusion

Results for the Processing step exhibits that most of the states are against Aadhaar or have complaints regarding its implementation. More disagreement is presented from developed states i.e. 52.09% of the total, as compared to the underdeveloped states with 47.01%. This further ensures the fact that Aadhaar is helping more people in underdeveloped states whereas people in the developed states are unhappy with the need to mandate it for various services.

The result of the sentiment analysis reveals that 13(40.62%) states were for Aadhaar and 19(59.38%) were against it. 10 out of 11 (90.9%) developed states had negative views whereas only 9 out of 21 Underdeveloped sates/UT showed disagreement towards it (42%). Even in recent days the government has demanded to make Aadhaar mandatory for day to day needs such as compensation to buy ration for citizens below the poverty line or pensions for the elderly, this has caused discouragement from the side of the common people as well as the Supreme Court. If the government would give it some more time to be adapted by the people and improve its security protocols then it would face lesser opposition from the general public.

References

[1] 600 million citizens of India are now enrolled with biometric ID- John Daugman (May 2014)
[2] About UIDAI: Official UIDAI website by the Government of India. (2018)
[3] Lexicon-Based Methods for Sentiment Analysis MaiteTaboada Simon Fraser University Julian Brooke University of Toronto Milan Tofiloski Simon Fraser University Kimberly VollUniversity of British Columbia Manfred Stede University of Potsdam (2011
[4] Sentiment Analysis of Twitter Data Apoorv Agarwal BoyiXie Ilia Vovsha Owen Rambow Rebecca Passonnea (June 2011)
[5] Twitter API – Up and running – Kevin Makice (March 2009)
[6] UIDAI Finance and Budget section. (2017)
[7] Assess the awareness level of people on Aadhaar Card as public wellbeing - Nimesh P. BhojakResearcher , Hemchandracharya North Gujarat University, India
[8] Digital India –A Unique Identity - Jayaprabha M. &Felcy Judith
[9] The Tweets They Are a-Changin':Evolution of Twitter Users and Behavior- YabingLiu, ChloeKliman-Silver, AlanMislove
[10] https://www.google.co.in/amp/www.thehindu.com/news/national/aadhaar-centres-onlyfrom-september/article19197918.ece/amp