GENERATING AN IMPACT ANALYSIS ALGORITHM FOR THE TOURISM BRAND IncredibleIndia

B.VeeraJyothi

CBIT,India

B.Aishwarya

CBIT,India

S.Guhapriya

CBIT,India

Abstract - **The Ministry of Tourism undertakes publicity campaigns in the Print, TV, Digital media, etc. for visibility of brand Incredible India, to create awareness about India among overseas travellers and to encourage Indians to explore their country. A methodology is required to analyse the effectiveness of such campaigns being undertaken by the Ministry of Tourism, both in the domestic and international media. There are three parts discussed in this paper: First, we aim to collect data (both from National and International media, online and Social Media) and create a sample dataset and analyse the impact being generated by such ads. Secondly, we generate a generic algorithm that can deal with any such compiled data set and returns the impact – Nationally and Internationally – in certain metric terms. Thirdly, we aim to provide an analysis of where ads related to the brand Incredible India, may fare better (get improved responses or where responses can be recorded), when posted. This is a real-time problem being faced by the Ministry of Tourism of India, when trying to promote the tourism brand of our country – Incredible India!**

*Index Terms -***Incredible India, Ministry of Tourism, Twitter API**

I INTRODUCTION

Tourism in India is economically important and is growing rapidly. The World Travel & Tourism Council calculated that tourism generated ₹14.02 lakh crore (US$220 billion) or 9.6% of the nation's GDP in 2016 and supported 40.343 million jobs, 9.3% of its total employment. The sector is predicted to grow at an annual rate of 6.8% to ₹28.49 lakh crore by 2027 (10% of GDP).In October 2015, India's medical tourism sector was estimated to be worth US$3 billion. It is projected to grow to $7–8 billion by 2020. In 2014, 184,298 foreign patients traveled to India to seek medical treatment. The Indian Government, every year, dishes out crores of Rupees in order to advertise and put forth India as the ultimate tourist destination. With incredible diversity present in the expanses of this country, India has great potentials of turning into one of the best tourist destinations of the world. In order to make sure that the investments made by the government are utilized in the optimum manner, we devise to come up with an algorithm. We aim at building an algorithm that works towards fulfilling the following

- Lower investment costs.
- Save resources.
- Accurately predict how to improve adverts.

The problem statement we are dealing with in this paper is, The Ministry of Tourism undertakes publicity campaigns in the Print, TV, Digital media, etc. for visibility of brand **Incredible India**, to create awareness about India among overseas travelers and to encourage Indians to explore their country. A methodology is required to analyze the effectiveness of such campaigns being undertaken by the Ministry of Tourism, both in the domestic and international media. The main focus of our paper lies on "Impact Analysis." We aim at coming up with a generic algorithm that can easily be scaled to serve different media outlets. The input for this generalized algorithm is a generalized data set. The means of developing this data set varies with different kinds of media outlets. The output generated by this algorithm is a histogram that aides in visualization of the results along with a detailed analysis that suggests the next steps to be taken. The output generated thus aims at analyzing the REACH, REVIEW and BUSINESS INTELLIGENCE of the adverts and campaigns taken up by the government.

The steps involved in this project, in general terms are,
- Simple data collection takes places from our chosen domain
- Through sentimental analysis of the data, we obtain the positive, neutral and negative opinion percentage.
- Impact analysis of the parameters provided by Step 2 is done, to provide graphical results.

There are various steps involved in this project, which in turn require the assistance and knowledge in various domains. The different domains involved in this project are,
- Text Mining.
- Sentimental Analysis.
- Data Visualization and Text summarization (for analyzing impact).

Fig 1. Implementation Steps

II MOTIVATION

According to the World Travel and Tourism Council, Indian tourism revenue grew 6.9 percent to almost $42 billion in 2010 despite a global recession. This compares to a worldwide tourism industry that remained flat during the same time period. Domestic tourism fueled much of the growth as Indians increasingly visited other areas of their country, but international arrivals also played an important part in the industry's expansion. Just over five million international tourists arrive in India annually (2009), and while that represents less than 1 percent of total worldwide international travel, tourists to India spend almost twice as much as the average international tourist.Tourism, benefits India in three ways---employment, foreign currency and infrastructure development. Over 20 million people work in India's tourist industry. This includes jobs in hotels, transport, attractions and tour companies. Others indirectly benefit such as restaurants and retail shops. Tourism generates over $11 billion in foreign currency, assisting the country's trade balance. Another benefit is infrastructure development such as hotel construction, airport improvements and ground transportation systems. As an example, the government set a deadline for opening a major expansion of Delhi's metro system concurrent with the opening of the 2010 Commonwealth Games. The city met the deadline and welcomed thousands of tourists with improved transportation throughout the city. In addition, tourism promotes understanding among people, both among Indians from different parts of the country, and with foreigners.Despite this, a lot of the investments made fail to reach the targeted audience. Since the 21st century, the nation has been continuously failing to improve the number of both national and international tourists. Although there is some significant increase in the number of domestic tourists, the number of international tourists coming to India has remained stagnant. The graph below shows the same. While the number of domestic tourists has been undergoing a steady increase, there is still a lot of room for improvement.
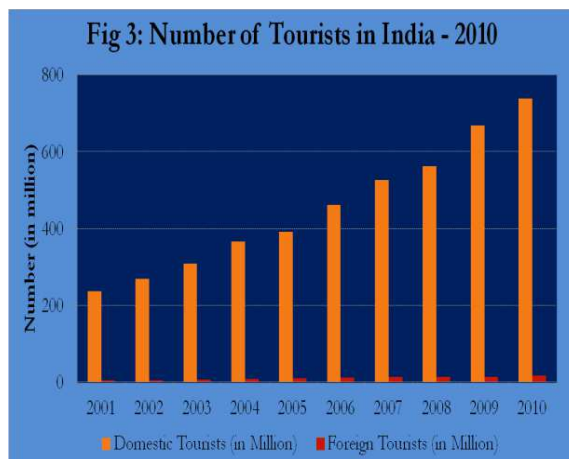

Fig 2. Tourism Statistics

III PRESENTLY USED MODELS

As we have already discussed, this paper is a combination of three different domains that contribute towards each step of this process.The first step involves Text Mining. This step depends a lot on the kind of media outlets we are going to use in order to mine the data. We need to mine the data from the targeted source. There are several methods that already exist in order to do this. For instance, if we are looking to mine data from a social networking site, we need to use the API to get the tweets or the posts. Once the data from the specific sites are gathered, various kinds of text mining techniques like, Clustering, Information Extraction, Topic Tracking, Summarization, Categorization, Concept Linkage, Topic Visualization and Questions and Answers are used in order to generate proper inputs for the algorithm. The second step involved Sentiment analysis. There are many ways of doing this. First, after obtaining the data set, we need to find out the overall sentiment of the document. We can also rate the document as positive, negative or neutral in terms of various aspects present in the document. The entire process can either be done manually or by using different sentiment analysis tools that already exist. There exist tools like IBM Watson, Tweetstats, WordCloud, etc.

The third step, involves data visualization and text summarization. The tools that can perform this are, Google charts, Highcharts, Leaflet, Timeline JS etc. After performing data visualization, we need to perform analysis and text summarization. There are many existing tools and models that can perform the task. They include CSVkit, DataTables, FreeDive, etc.A model that does all of the mentioned steps is employed in a lot of domains. We aim at producing this model in order to improve the advertisements for Incredible India. This is a real time problem and hence, there are little to none pre-existing models that perform these steps.

IV PROPOSED MODEL

Social media has become one of the most important sources of information regarding the opinions of people. It is a significant factor in order to study the market and reactions of various sections of people.

Hence, in this paper, we intend on mining and performing analysis of people's opinions on Twitter, one of the most popular social networking sites.

1. We are first going to concentrate on how to mine data from twitter. In order to do this, we need to first obtain the key and id for the API. Once this is done, we can derive all the necessary data, along with various kinds of side information (like shares, retweets and likes) to aid our process. Once all the data is obtained, we are supposed to introduce uniformity in the data in order to make it more generic.

2. After the generic data is obtained, we are going to feed it in the algorithm and perform sentimental analysis. We are going to use 'WordCloud' for this. This tool extracts the top 10 most negative tweets and top 10 most positive tweets and presents it as the output.

![IJEEE logo]

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 5, Issue 3, March 2018)

3. Once the output from the sentiment analyzer has been obtained, we have to feed it into a data visualization tool and turn the data into required histograms. Using this histogram, we are going to present further analysis that will present the user with pointers on how they can optimize the way advertisements are made.
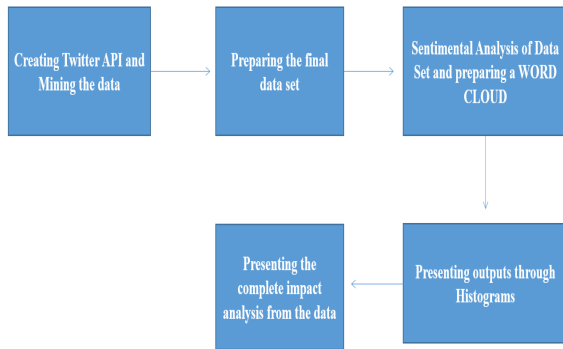

Fig 3. ARCHITECTURE (An Example of the Twitter Process)

V IMPLEMENTATION

SOFTWARES USED: R Tool, Twitter API and Tableu Public.
Step 1: Using the Twitter API (User ID, API Key), which provides us with the data to manufacture our data set which consists of – *(USER_AGE::TWEET::SHARES_RETWEETS::EMOTICON_ SCORE::EYEBALLS_PER_HOUR::TIMING_CONSISTENCY).*(Source-#IncredibleIndia Tweets on Republic Day, 2018).
Step 2: Once, we have compiled the data set, we begin the algorithm construction by putting this data set as inputs for the Algorithm.
Step 3: The Algorithm begins with the sentimental analysis of the given data set using Word Clouds. From these Word Clouds, we not only get the Negative, Positive and Neutral words, but we also get the Top 10 positive words from the WordCloud.
Step 4: **Normalization** – *EMOTICON SCORE NORMALIZATION(Using AVERAGE ABSOLUTE DEVIATION)* (Integrating Emoticon score with the Polarities – Twitter API provides an Emoticon score that needs to be integrated with the Polarity range of –2 to +2)
POLARITY RANGE NORMALIZATION (Using MIN-MAX Normalization in R TOOL) (To conform the polarity range for maximum accuracy – Once the Emoticon score has been normalized, there may be deviations in the polarity range. Therefore, Min max normalization is done to conform it to the proper range of –2 to +2).
From the word cloud, the sentimental analysis results are then passed on to the connected R Tool where the Business Intelligence algorithm is applied to provide the following analysis in the form of Histograms as the data visualization tool –

TWITTER:
Polarities vs Number of Tweets
Polarities vs Top 10 Aspects
Polarities vs Age
Step 5: Once these Histograms are displayed further analysis is done by providing an integrated analysis of the results from both the offline and online media portals (Social Media).
Step 6: To show the international brand image of Incredible India, we also provide the geographical scattering or the origins of the tweets collected under that hashtag on Twitter, through a business intelligence tool, called Tableu Public.

VI RESULTS

1. **Pre-Processing Data** (Removing Retweets, Twitter IDs, Separating Emoticons with score, Special Characters, Columns with irrelevant regions)
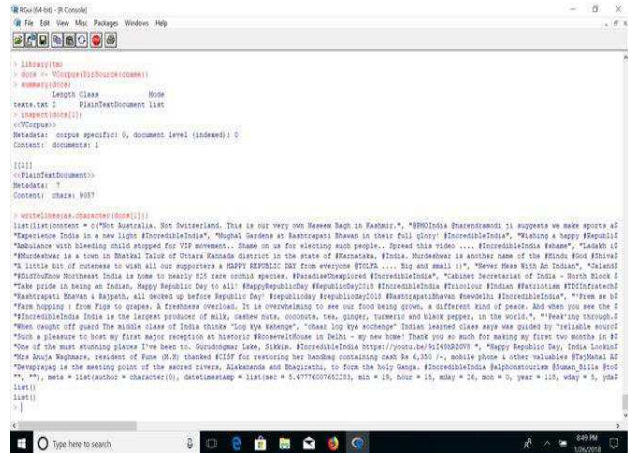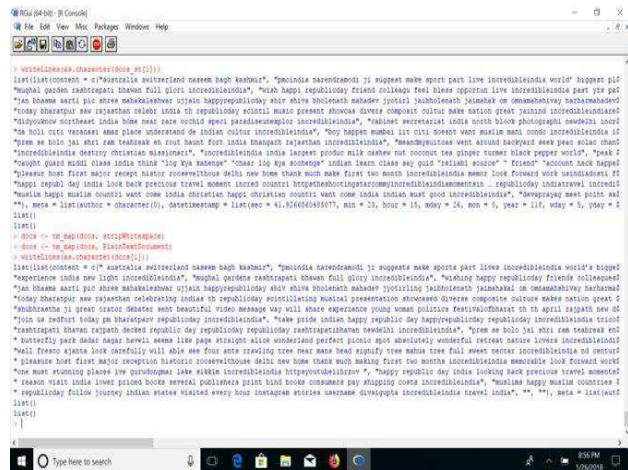

Fig 4.Loading Data For Pre-processing

Fig 5.Pre-processing Data

A. Generating WordClouds: (Twitter)

Fig 6. Twitter Data Set

Fig 7. Loading Data IntoWordCloud

Fig 8. Twitter Based Term Doc Matrix

Fig 9. Twitter Based WordCloud

Fig 10.Emoji Score Normalization

Fig 11. Data Set After Emoji Score Normalization

Fig 12. Average Absolute Deviation (Integrating Emoji Score With Polarity Range)

Fig 13.Polarity Range Normalization:

![IJEEE logo]

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 5, Issue 3, March 2018)

Min-Max Normalization

Above data frame could be normalized using Min-Max normalization technique which specifies the following formula to be applied to each value of features to be normalized. This technique is traditionally used with **K-Nearest Neighbors (KNN)** Classification problems.

```
1  (X - min(X))/(max(X) - min(X))
```

Above could be programmed as the following function in R:

```
1  normalize <- function(x) {
2      return ((x - min(x)) / (max(x) - min(x)))
3  }
```

In order to apply above normalize function on each of the features of above data frame, df, following code could be used. Pay attention to usage of **lapply function**.

```
1  dfNorm <- as.data.frame(lapply(df, normalize))
2  # One could also use sequence such as df[1:2]
3  dfNorm <- as.data.frame(lapply(df[1:2], normalize))
```

Fig 14. Min-Max Normalization

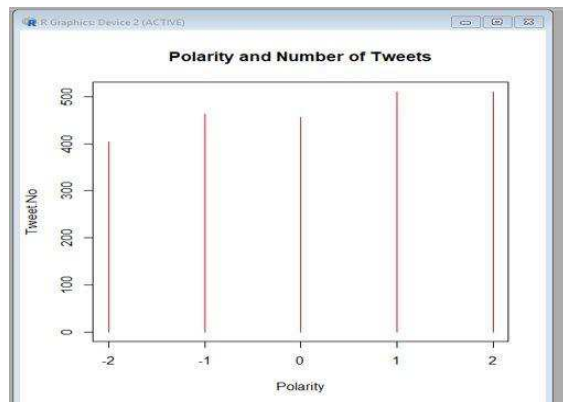| Tweet No	Tweets	Polarity
1	Not Australia. Not Switzerland. This is our very own Naseem Bagh in Kashmir.	0
2	@PMOIndia @narendramodi ji suggests we make sports a part of our lives	1
3	'backyard' of CM @tsrawatbjp for past 3 weeks. Get Outdoors #uttarakhand #healthychoices #healthyfamily	2
4	Catch sight of #Binsar in #Devabhoomi #Uttarakhand of #incredibleIndia It's a quaint hamlet & famous #HillStation known for it's marvellous vista of the #Himalayas ,virid #AlpineForests ,pacific vibes,rich flora & fauna & pleasant weather.	0
5	Experience India in a terrible light #incredibleIndia	-1
6	Mughal Gardens at Rashtrapati Bhawan is the most unclean place have ever been to! #incredibleIndia	-2
7	Wishing a happy #Republicday2018 to my friends & colleagues in Feeling blessed 2have had the opportunity 2live in #incredibleIndia for the past 1.5 yrs &be part of the amazing #GrowingPartnership	0
8	Not Australia. Not Switzerland. This is our very own Naseem Bagh in Kashmir.	0
9	@PMOIndia @narendramodi ji suggests we make sports a part of our lives	0
10	'backyard' of CM @tsrawatbjp for past 3 weeks. Get Outdoors #uttarakhand #healthychoices #healthyfamily	1
11	Catch sight of #Binsar in #Devabhoomi #Uttarakhand of #incredibleIndia It's a quaint hamlet & famous #HillStation known for it's marvellous vista of the #Himalayas ,virid #AlpineForests ,pacific vibes,rich flora & fauna & pleasant weather.	2
12	Experience India in a new light #incredibleIndia	2
13	Mughal Gardens at Rashtrapati Bhawan in their full glory! #incredibleIndia	2
14	Wishing a happy #Republicday2018 to my friends & colleagues in Feeling blessed 2have had the opportunity 2live in #incredibleIndia for the past 1.5 yrs &be part of the amazing #GrowingPartnership	2
15	Not Australia. Not Switzerland. This is our very own Naseem Bagh in Kashmir.	1
16	@PMOIndia @narendramodi ji suggests we make sports a part of our lives	1
17	'backyard' of CM @tsrawatbjp for past 3 weeks. Get Outdoors #uttarakhand #healthychoices #healthyfamily	1
18	Catch sight of #Binsar in #Devabhoomi #Uttarakhand of #incredibleIndia It's a quaint hamlet & famous #HillStation known for it's marvellous vista of the #Himalayas ,virid #AlpineForests ,pacific vibes,rich flora & fauna & pleasant weather.	1

Fig 15. Final Data Set

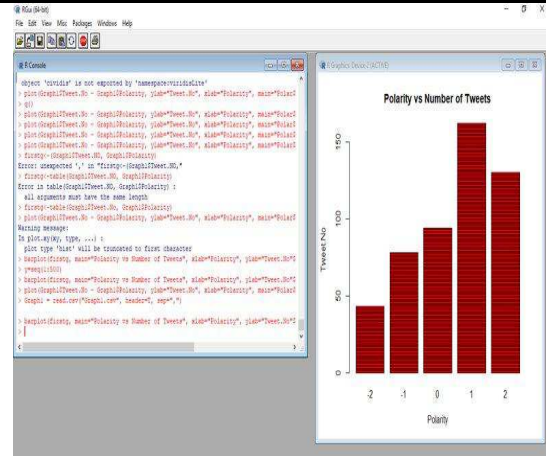Fig 16. Twitter (Polarities vs Number of Tweets) Part 1

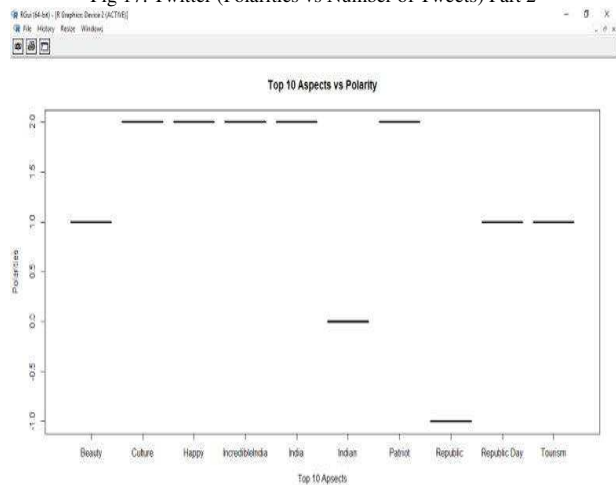Fig 17. Twitter (Polarities vs Number of Tweets) Part 2

Fig 18. Twitter (Polarities vs Top 10 Aspects)

Top 10 Aspects being (Obtained from WordCloud) - Beauty, Culture, Happy, IncredibleIndia, India, Indian, Patriot, Republic, RepublicDay, Tourism)
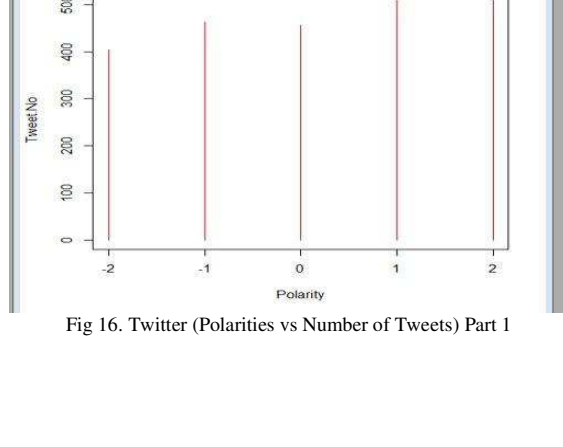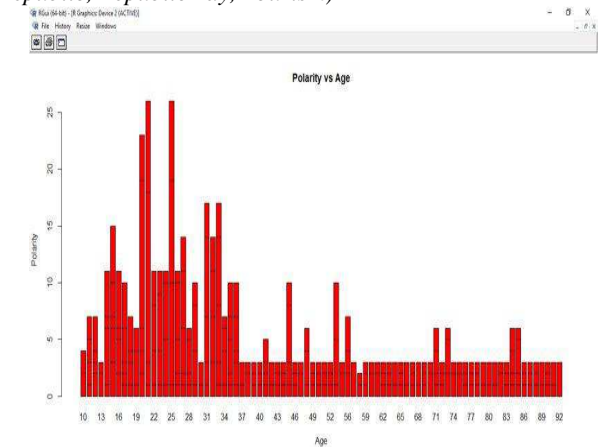

Fig 19. Twitter (Polarities vs Age)

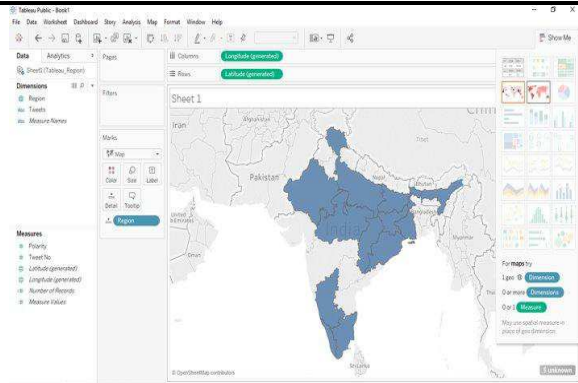B. Region Based Outputs for Twitter using Tableau Public -
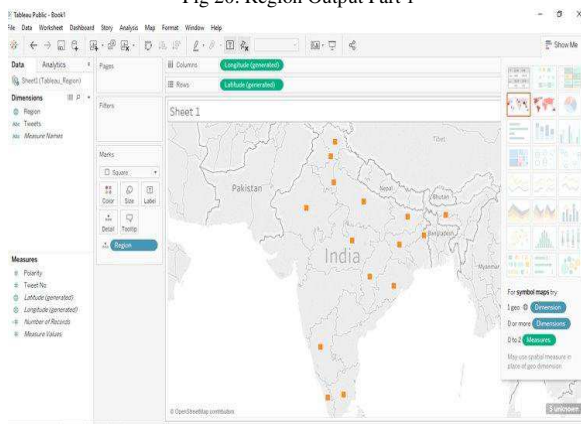
Fig 20. Region Output Part 1

Fig 21. Region Output Part 2

1. Optimizing advert tags on social media
2. Age group optimization to increase eyeballs
3. And finally presenting better advertisement propositions

REFERENCES

[1]. https://innovate.mygov.in/challenges/developing-an-algorithm-for-impact-analysis-of-various-publicity-campaign-being-undertaken-by-the-ministry-of-tourism-in-the-print-media-tv-online-and-social-media/
[2]. IEEE:http://ieeexplore.ieee.org/abstract/document/1509307/?reload=true
[3]. 'On the Use of Side Information for Mining Text Data' by Charu C. Aggarwal, Yuchen Zhao and Philip S. Yu, 2014.
[4]. 'A Survey Of Text Mining Techniques and Applications' by Vishal Gupta and Gurpreet S. Lehal, 2013.
[5]. 'Building System Rules and Algorithms' by Mike Shal, 2014.
[6]. 'An Empirical Comparison of Dynamic Impact Analysis Algorithm' Alessandro Orso, z TaweesupApiwattanapong,z James Law, Y Gregg Rothermel,y and Mary Jean Harrold, 2014.
[7]. https://www.computerworld.com/article/2502891/business-intelligence/business-intelligence-8-cool-tools-for-data-analysis-visualization-and-prese

ANALYSIS

TWITTER: -2 (405), -1 (469), 0 (461), 1 (502), 2 (501)
(Number of tweets under each polarity)
- Min Age = 10, Max Age = 92
- Highest Positive Response = Age group 35-41
- Lowest Negative Response = Age group 54-72

Impact Words – Culture, Happy, IncredibleIndia, India, Patriot

VII CONCLUSION

Our paper aims to comprehend and provide the basics of a Business Intelligence Algorithm that can analyse and generate the details about the given data set and what kind of information we can extract from it and how we can use it. The future scope of our particular algorithm can be improvised in many aspects that require cash resources to be invested before the information can be bought to use and analyse for the expected output.

- The future scope of the project is that it can be expanded to be applied to any customer other than Incredible_India in the future.
- The algorithm would return the impact analysis of their current adverts and provide suggestions to improve their advertisement target sector through: