WHAT IS TWITTER SENTIMENT ANALYSIS?

Twitter Sentiment Analysis is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention.

Hence Twitter is a public platform with a mine of public opinion of people all over the world and of all age categories.

WHY TWITTER SENTIMENT ANALYSIS?

Sentiment analysis is in demand because of its efficiency. Thousands of text documents can be processed for sentiment in seconds, compared to the hours it would take a team of people to manually complete. Because it is so efficient (and accurate – Semantic has 80% accuracy for English content) many businesses are adopting text and sentiment analysis and incorporating it into their processes.

PROBLEM STATEMENT

There is always need of statistical survey for any government or non-government organization to work efficiently. The four main reasons why the organizations should conduct survey are as follows:

1) To uncover the answer.
2) To evoke discussion.
3) Base decision on objective information.
4) To compare results.

The drawbacks of the traditional approach such as household surveys, census used for manual surveys, which are very expensive, time consuming and less effective.

As the usage of social media has been increased with a high rate, the data that is shared on social media by the users is analyzed to get the information about the people. Many applications of sentiment analysis have been implemented to know the different sentiments of the users. Our proposed work uses twitter data to analyze opinion of people pertaining to the proposed smart cities in India.

IMPLEMENTATION DETAILS

To implement the purposed system R 3.3.2 is used .The Implementation process consists of collection of twitter data pertaining to proposed smart cities and later performing sentimental analysis on it.

ROLE OF R

R is a language and environment for statistical computing and graphics. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:

1) An effective data handling and storage facility.
2) A suite of operators for calculations on arrays, in particular matrices.
3) A large, coherent, integrated collection of intermediate tools for data analysis.
4) Graphical facilities for data analysis and display either on-screen or on hardcopy, and
5) A well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

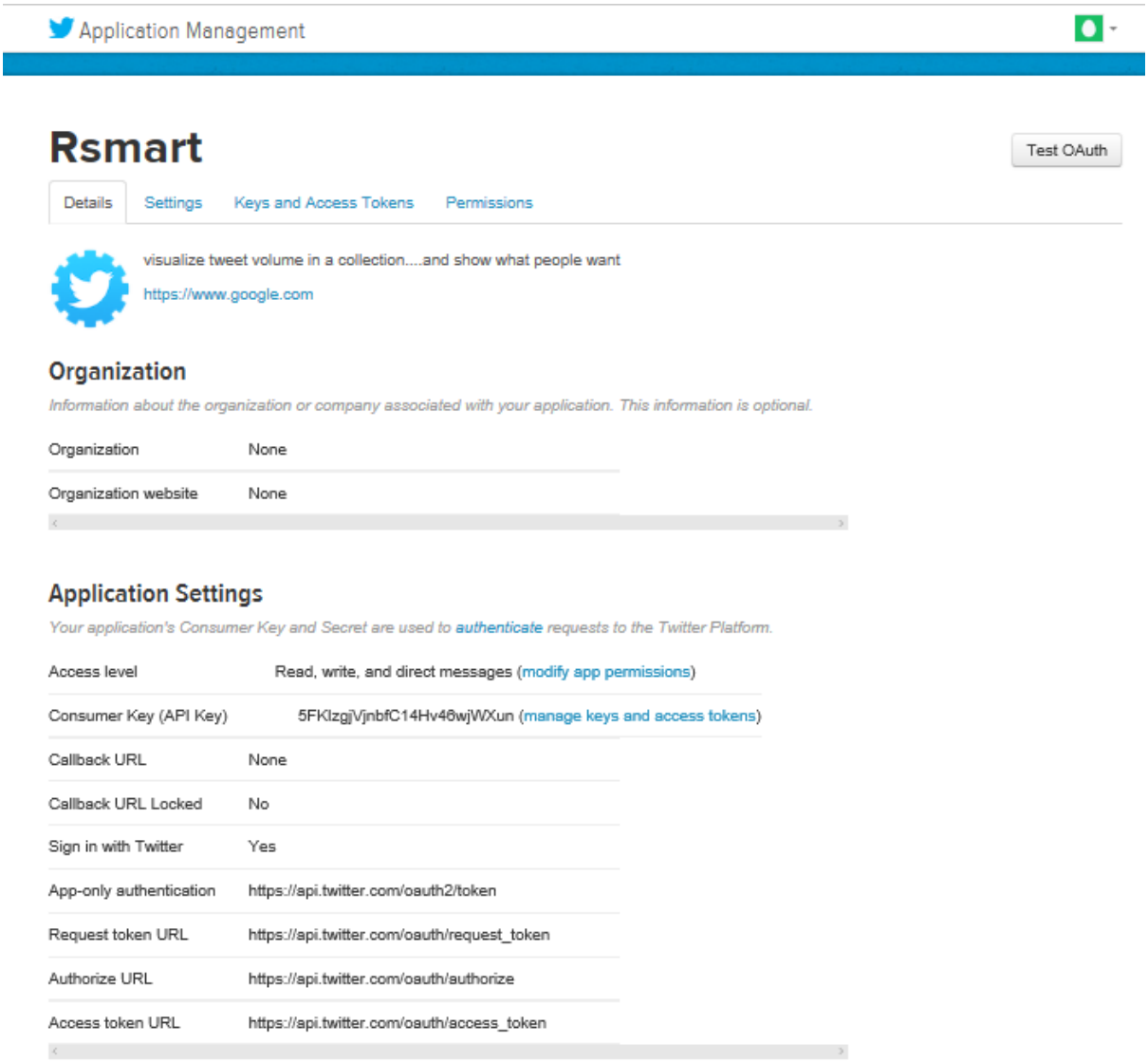
PACKAGES USED

- twitteR: Provides an interface to the Twitter web API
- **ROAuth**: Provides an interface to the OAuth 1.0 specification allowing users to authenticate via OAuth to the server of their choice.
- **RCurl**: Provides functions to allow one to compose general HTTP requests and provides convenient functions to fetch URIs, get & post forms, etc. and process the results returned by the Web server.
- **ggplot2**: An implementation of the grammar of graphics in R. It combines the advantages of both base and lattice graphics: conditioning and shared axes are handled automatically, and you can still build up a plot step by step from multiple data sources.
- **tm** : A framework for text mining applications within R.
- **wordcloud**: visual representation in the form of wordcloud where size of the word is proportional to the frequency of words used in the tweets.
- **plyr**: Tools for Splitting, Applying and Combining Data.
- **stringr:** stringr is a set of simple wrappers that make R's string functions more consistent, simpler and easier to use.

CREATING TWITTER APPLICATION

First step to perform Twitter Analysis is to create a twitter application. This application will allow us to perform analysis by connecting our R console to the twitter using the Twitter API. The steps for creating our twitter applications are:

- First we go to https://dev.twitter.com and login by using our twitter account.
- Then we go to My Applications and Create a new application.
- We give our application a name, describe about application in few words, provide website's URL or blog address. Leave the Callback URL blank for now. Complete other formalities and create our twitter application. Once, all the steps are done, the created application will show as below and note the Consumer key and Consumer Secret numbers as they will be used in RStudio later.



SAVING TWEETS

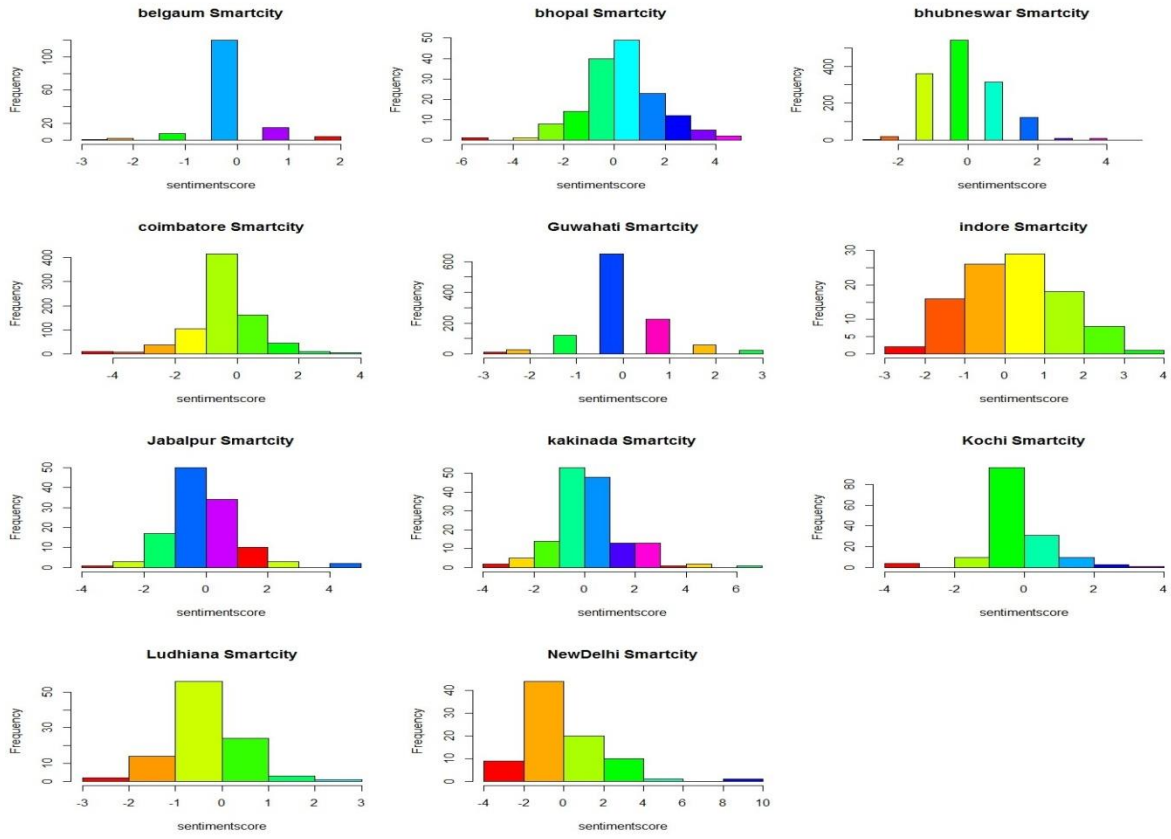
Once the authentication is done and authorized by twitter, we can fetch most recent tweets related to any keyword. I have used #smartcityindia, #smartcitybhopal, #smartcitybhubneswar, #smartcitychennai, #belgaumsmartcity, #smartcitycoimbatore, etc as proposed smart cities of India is the most talked about in India now a day.

CLEANING TWEETS

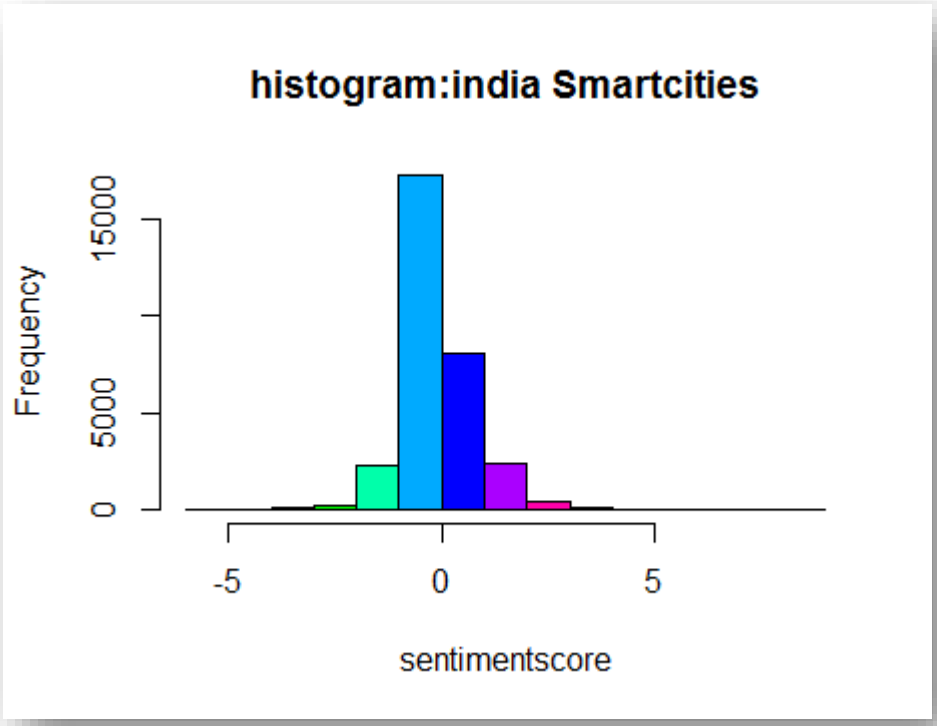
The tweets are cleaned in R by removing:

- Extra punctuation
- Stop words (Most commonly used words in a language like the, is, at, which, and on.)
- Redundant Blank spaces
- Emoticons
- URLS

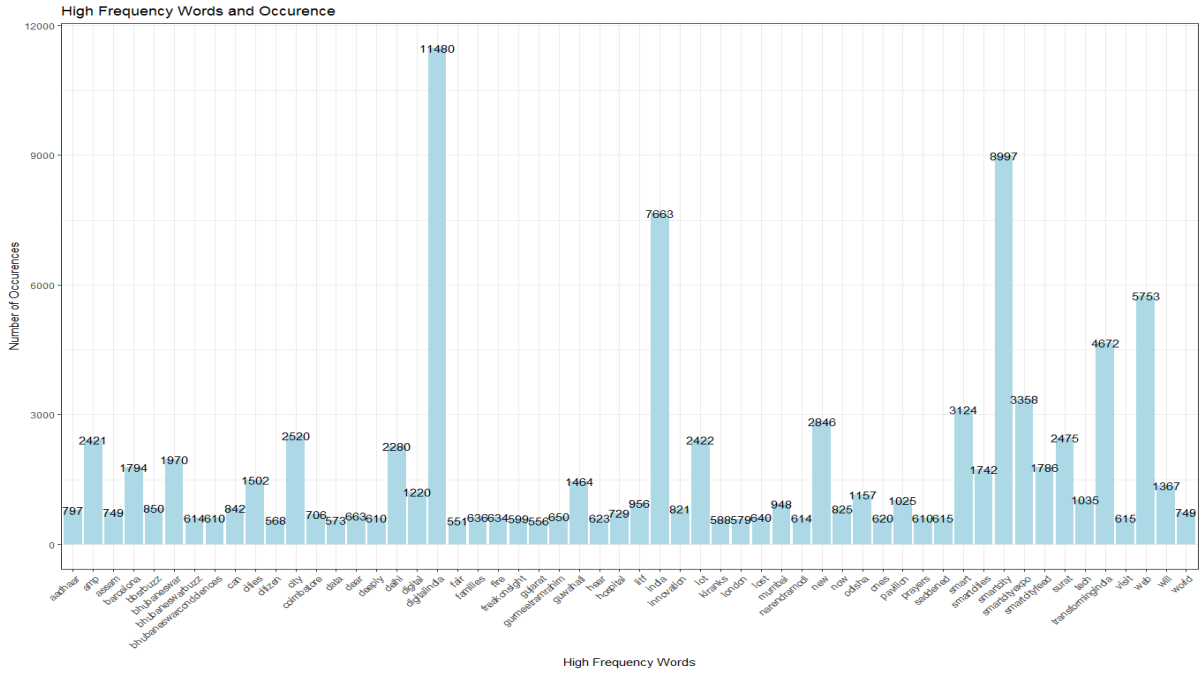
VISUALIZATION OF VARIOUS SMART CITIES



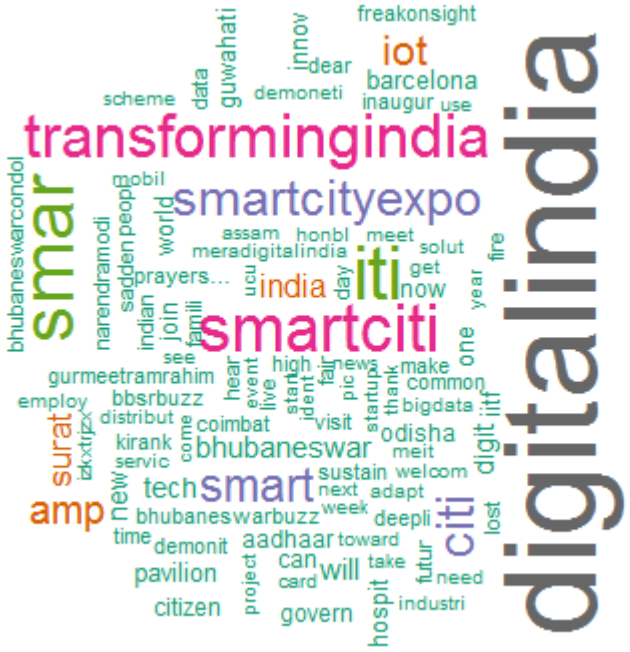
COMBINED VISUALIZATION OF SMARTCITIES



MOST FREQUENT WORDS



WORD CLOUD


digitalindia
transformingindia
smartcityexpo
smartciti
smar
iti
iot
barcelona
freakonsight
guwahati
innov
dear
demoneti
inaugur
scheme
data
use
mobil
world
assam
honbl
meet
solut
meradigitalindia
india
day
get
now
year
fire
prayers...
ucu
sadden
peopl
narendramodi
indian
join
famili
bhubaneswarcondol
see
one
gurmeetramrahim
hear
event
live
high
start
ident
fair
pic
startup
thank
make
common
employ
bbsrbuzz
bigdata
itif
distribut
coimbat
visit
odisha
digit
surat
kirank
servic
come
bhubaneswar
meit
new
tech
smart
sustain
welcom
next
adapt
amp
bhubaneswarbuzz
week
deepli
citi
lost
time
demonit
aadhaar
toward
pavilion
project
can
card
will
take
futur
need
citizen
govern
hospit
industri