

PROJECT

---

# **TWITTER MINING**

using R language

---

March 24, 2017

Name: Soumya Ranjan Patel

Registration ID: 14010203

Veer Surendra Sai University of Technology, Burla  
Department of Computer Science and Engineering

## INTRODUCTION

Big Data is a rapidly expanding technical sector. The main priority of this project is to get a good baseline understanding in Big Data. BIG DATA is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. Big Data is the technical term that describes the availability and growth of both structured and unstructured data.

This project is an attempt to understand the baseline of Data Processing and Semantic analysis of the data collected from the Twitter on basically two topics.

I. USA Election-2016

II. Note-ban in India (2016)

About 1000 tweets were collected and stored, called corpus. These data were then analysed using the libraries. The output data were shown in image which can be converted to either portable document format (pdf) or the various image format (viz. jpf, png etc).

## APPROACH

In this project the big bulk of data is analysed in the following ways.

I. I got registered in the Twitter developer account.

II. The tweets were collected using key works and access tokens.

III. The sentiment package was installed for 'R'.

IV. The data analysis and sentiment analysis were done.

## TOOLS USED

In this project the following tools used were:

I. R Studio Version 3.0

II. sentiment library.

III. ggplot library.

IV. wordcloud library.

## **REFERENCES**

- I. Twitter Mining Tutorials.
- II. Various sources on the Internet.