



विश्वजीवनमृतं ज्ञानम्

Sentiment Analysis of Tweets and News Media Reporting on Demonetization Issues

Problem Statement Draft for
B.Tech Project

Submitted by

Group No: 24

Ambuj Mishra (2014IPG-012)

Sheshan Sheniwal (2014IPG-080)

Sunil Kumar (2014IPG-117)

Under the guidance of

Dr. Pradip Swarnakar

Objective:

Comparison of sentiments of different organisations on a particular topic “demonetization”

To find out the relationship between Online Tweets and the Newspaper articles on the topic and investigate through time that how those particular patterns have changed.

Background:

Generally, data mining is the process of analysing data from different perspectives and summarizing it into useful information.

There is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyse this huge amount of data and extract useful information from it.

When it is over we would be able to use this information in many applications such as Market Analysis, Fraud Detection, Production Control, Movie Reviews, Customer satisfaction, Product review etc.

Today Data mining became the need of time.

Motivation:

Everything is getting online now a days, be it trending news or entertainment, education or social life. In today's life people are much involved in the online news rather than one. Large number of people are talking on the same topic with different views and ideas, results in more information availability.

Because of the large amount of online information, it become strenuous to uncover what we actually require. For understanding, searching, organizing and automatically summarizing electronic annals, topic modelling plays an important role.

Problem:

The main problem of social networks is the arrangement of online statistics or data. Online data that is available is beyond limitations and need a good Organization schema.

Large amount of data is not easy to understand. To accomplish this task, we need to find out effective computational methods for searching and organizing enormous information. Nowadays people look at the online information frequently rather than offline Newspaper and they go to the main tools like search and links.

People suffer from following things:

- Can't find the actual content in first go
- Take much time to figure out what they need
- Human power is not enough to manage large collection of document

We shall arrange and get tweets and news articles sources, further summarize them, prepare RTF data in the database for algorithm implementation, extract embedded topics from collection of data, discover the hidden themes that pervade the tweets and articles, annotate the documents according to those themes, use annotations to organize, summarize, compare and search the texts and test them under varied time-intervals, authors, publications etc.

Timeline

Task 1: Include the basic study of data mining and understanding the problems through research papers.

Task 2: Collecting data (Tweets, News article, Rtf format)

2.1: Separating the data based on its title, date, words, and media type.

2.2: Create a database of direct speech of the reports and mark the name of the individual and organizational name.

2.3: Saving all the extracted documents into final database.

Task 3: Sentiment analysis with at least two existing APIs

3.1: Summarize the data.

3.2: Use sentiment analysis tools to get positive or negative feedback of demonetization.

Task 4: Analysis and result report preparation.

S. No.	Task	Days	Completion date
1	Studying research papers	5	13/06/2017
2	Data gathering	7	20/06/2017
3	Data extraction	20	10/07/2017
4	Data storing into database	5	15/07/2017
5	Sentiment analysis	7	21/07/2017
6	Analysis and result	5	26/07/2017
7	Report preparation	7	03/08/2017

Gantt chart

Tasks



Analysis and
result

Report
preparation

Weeks

W1 W2 W3 W4 W5 W6 W7 W8 W9

