****

**SYNOPSIS OF MAJOR PROJECT**

**Date :** 11th January, 2018

**Major Project Title :** Budget 2018 Analysis using Tweets

**Name of Guide :** Mrs. Chitra Nasa

|  |  |  |  |
| --- | --- | --- | --- |
| Programme :- **B.Tech CSE** | | Year/Semester :- **8th Semester** | |
| S.No | Enrollment No. | Name | Signature |
| 1) | 11813302714 | Shivani Singh |  |
| 2) | 12813302714 | Divyanka Latwal |  |
| 3) | 30413302714 | Karan Dawar |  |
| 4) | 60413302714 | Surabhi Nanda |  |

**Major Project Summary :**

**Resource Requirement :**

**Schedule of Major Project work Completion :**

Signature of Student(s) Signature of Guide Signature of Proctor

**Approval by Board of Faculty**

|  |  |  |
| --- | --- | --- |
| Member | Signature | Remark(Approved/Not Approved) |
|  |  |  |

**Content**

1. Abstract
2. Introduction
3. Objective Of Project
4. Technology Used
5. Reference

**Abstract**

With the advancement of web technology and its growth, there is a huge volume of data present in the web for internet users and a lot of data is generated too. Internet has become a platform for online learning, exchanging ideas and sharing opinions. Social networking sites like Twitter, Facebook, Google+ are rapidly gaining popularity as they allow people to share and express their views about topics, have discussion with different communities, or post messages across the world. Twitter, one of the largest social media site receives tweets in millions every day. This huge amount of raw data can be used for industrial or business purpose by organizing according to our requirement and processing. This project provides a way of Budget 2018 analysis using Tweets. Analyzing this vast quantity of unstructured data presents challenges for software and hardware.

**Introduction**

Nowadays, the age of Internet has changed the way people express their views, opinions. It is now mainly done through blog posts, online forums, product review websites, social media ,etc. Nowadays, millions of people are using social network sites like Facebook, Twitter, Google Plus, etc. to express their emotions, opinion and share views about their daily lives. Through the online communities, we get an interactive media where consumers inform and influence others through forums. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. People mostly depend upon user generated content over online to a great extent for decision making. For e.g. if someone wants to buy a product or wants to use any service, then they firstly look up its reviews online, discuss about it on social media before taking a decision. The amount of content generated by users is too vast for a normal user to analyze. So there is a need to automate this, various sentiment analysis techniques are widely used. Sentiment analysis (SA)tells user whether the information about the product is satisfactory or not before they buy it. Marketers and firms use this analysis data to understand about their products or services in such a way that it can be offered as per the user‟s requirements. Textual Information retrieval techniques mainly focus on processing, searching or analyzing the factual data present. Facts have an objective component but,there are some other textual contents which express subjective characteristics. These contents are mainly opinions, sentiments, appraisals, attitudes, and emotions, which form the core of Sentiment Analysis (SA). It offers many challenging opportunities to develop new applications, mainly due to the huge growth of available information on online sources like blogs and social networks.

Objective of Project

This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream. To build a model that can analyses the response of people on particular topic.

Technology Used

R Language :

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S.

Shiny Server (Use for Deployment of Project / For GUI Purpose) :

Shiny Server is a back end program that makes a big difference. It builds a web server specifically designed to host Shiny apps. ... You can also use Shiny Server to make your apps available across the Internet when you choose.

**References**

1. Yanchang Zhao. R and Data Mining: Examples and Case Studies. ISBN 978-0-12-396963-7, December 2012. AcademicPress, Elsevier. 256 pages.
2. <http://www.rdatamining.com/resources/courses>
3. <http://www.rdatamining.com/resources/onlinedocs>
4. <https://www.toptal.com/python/twitter-data-mining-using-python>