SENTIMENTAL ANALYSIS OF TWEETS USING R

Sentimental analysis is defined as the process using which the given data can be determined whether it is positive, negative or neutral. It is commonly referred to as opinion mining as it is used to derive the opinion or attitude of the speaker. Twitter is a popular form of social media where people from across the globe present the views or opinions about various topics. When many people tweet about a common topic, the topic is considered to be trending. The tweets however, may have a positive, negative or a neutral opinion about the topic. R language is very versatile and has many packages available for the benefit of users. These packages can be used to perform the three stages of developing this project. The stages can be classified as:

i) Data Mining: Extracting Data from Twitter
ii) Sentimental Analysis of Mined Data
iii) Visual Representation of the Result of Analysis

Primarily, packages such as ROAuth, twitter, ggplot2 must be installed in the R environment.

Phase 1: A Twitter application must be created and valid API keys and access token keys must be obtained. These in combination of appropriate URLs can be used to establish a connection with the Twitter API. Tweets can be mined using the commands offered by the 'twitteR' and these tweets can be stored as a dataset on which analysis can be performed. Tweets can be extracted based on trending topics (eg. #GST) or by providing the twitter handle of a particular user's account (eg. @PMOIndia). The number of tweets to be extracted can also be specified. The location from where tweets are to be extracted can also be specified using WOEID (Where On Earth IDentifier).

Phase 2: The mined data is now available for sentimental analysis. The dataset is cleaned of any unnecessary characters such as punctuations, control characters, and digits. This cleaned data is converted into a vector of text to score the tweets. Text files containing positive and negative words are used to perform string matching of the cleaned dataset and in the process the tweets are scored such that a point is awarded for each positive word matched and a point is deducted for each negative word matched. If the final score is above zero, the tweet is considered to be of positive sentiment. Else, if the final score is less than zero, the tweet is considered to be of negative sentiment. Otherwise, the tweet is said to of neutral sentiment.

Phase 3: The result of the analysis of the dataset can be represented visually using the 'ggplot2' package. The polarity categories are mainly classified as positive, negative, and neutral. The number of tweets are represented on one axis and the polar categories on the other axis. Accordingly, the type of graph can be specified and the results of sentimental analysis of the tweets extracted can be represented.

Applications: Sentimental Analysis of tweets allows users to gain a better understanding of public opinion on certain topics. The results of this analysis can be used to make informed decisions in various fields. From determining the response of customers to a particular product to realizing the popular outlook of public on a particular topic, the applications of sentimental analysis are broad and powerful.