## 4.1 Project Introduction

## NAME: Twitter Sentimental Data Analysis

As we all know there are various types of data users on twitter. Every user has own views about anything. My project was all about analysing sentiments of twitter users by using set of commands accordingly. The main perspective of Sentiment Analysis refers to the use of text analysis to identify and extract subjective information in textual contents. There are two type of user-generated content available on the web – facts and opinions. Facts are statements about topics and in the current scenario, easily collectible from the Internet using search engines that index documents based on topic keywords. Opinions are user specific statement exhibiting positive or negative sentiments about a certain topic. Generally, opinions are hard to categorize using keywords. Various text analysis and machine learning techniques are used to mine opinions from a document. Sentiment Analysis finds its application in a variety of domains.

## 4.2 Benefit of Project

The main function of this project is to collect tweets or feeds from twitter as data in order to determine Positive and Negative percentage and viewed a visualization based on twitter data.

It can also be used by various companies to find out the demand and need of some specific project.

Sentimental analysis is also useful in determining the views of people on a particular political issue.

**4.3 Steps to Develop Project**

The first step was to fetch Twitter’s using a developers account. We can only take data from Twitter by developers account. Developers account is first verified by twitter, then only we get access to Twitter’s data. On twitter data are combination of both structured and un-structured data. Data can be in form of audio, video, text etc.

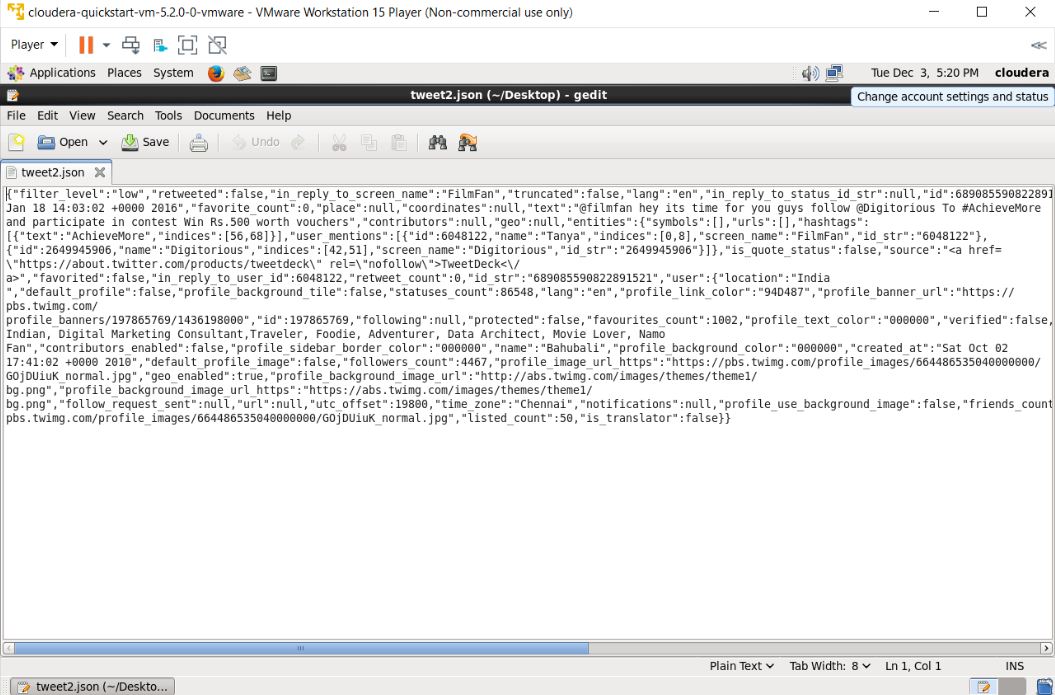


Fig 4.3.1: Screenshot of Data taken from Twitter

After taking the data I convert the data into a tabular format or we can say stored it in a table as data in tabular is easy to understand and represent as compared to data in normal data representation in paragraph form.

**Queries**

CREATE DATABASE PPT

CREATE EXTERNAL TABLE tweets(id int, name char, favourite BOOLEAN, text String)

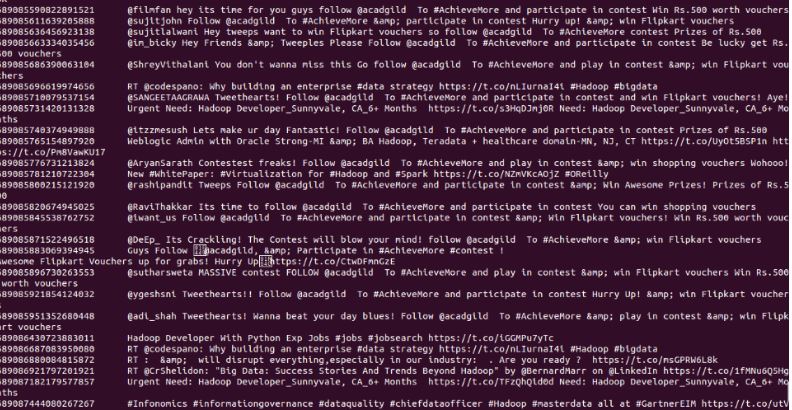
ROW FORMAT DELIMITED

FIELD TERMINATED BY (‘,’);

select \* from tweets

create table new\_words as select id as id,split(text,' ') as words from tweets;

select \* from new\_words;



create table dictionary(word string, rating int) ROW FORMAT DELIMITED FIELDS TERMINATED BY(‘,’);

LOAD DATA INPATH ‘all.txt' overwrite into TABLE dictionary;

select \* from dictionary

