**SENTIMENT ANAYLSIS Using Pig and R on Twitter**

library(twitter)

setup\_twitter\_oauth(consumer\_key =’’,consumer\_secret =’’,access\_token =

‘’,access\_secret =’’)

// IF doenst connect try multiple times

airData1=searchTwitter(searchString = "@airindiain",n = 5000)

d=twtodf(airData1)

da=d[c(id,text)]

write.table(da, file = "AD1.txt",row.names=FALSE, na="",col.names=FALSE, sep=",")

//Now put this file in HDFS.

hdfs.put(‘AD1.txt’,’./data/AD1.csv’)

//Now put AFINN.txt file in HDFS as well.

Hdfs.put(‘AFINN.txt’,’./data/AFINN.txt’)

//NOW Open PIG

// REGISTER Piggybank jar file

REGISTER /usr/lib/pig/piggybank.jar;

extract\_details = LOAD 'hdfs://127.0.0.1:8020/user/cloudera/data/AD1.csv' using org.apache.pig.piggybank.storage.CSVLoader() AS (id:long,text:chararray);

tokens = foreach extract\_details generate id,text, FLATTEN(TOKENIZE(text)) As word;

dictionary = load '/AFINN.txt' using PigStorage('\t') AS(word:chararray,rating:int);

word\_rating = join tokens by word left outer, dictionary by word using 'replicated';

describe word\_rating;

rating = foreach word\_rating generate tokens::id as id,tokens::text as text, dictionary::rating as rate;

word\_group = group rating by (id,text);

avg\_rate = foreach word\_group generate group, AVG(rating.rate) as tweet\_rating;

positive\_tweets = filter avg\_rate by tweet\_rating>=0;

DUMP positive\_tweetssw

negative\_tweets = filter avg\_rate by tweet\_rating>=0;

DUMP negative\_tweets

//Or you could use

STORE avg\_rate INTO ‘SentimentAnalysis.txt’ using PigStorage(',');

// By Storing the data like this you will be able to do other operations on analysed data.