SESSION 22 ASSIGNMENT 2

PROBLEM : Sentiment analysis on demonetization

Let us find out the views of different people on the demonetization by analysing the tweets from twitter. Here is the dataset where twitter tweets are gathered in CSV format. You can download the dataset from the below link

<https://drive.google.com/open?id=0ByJLBTmJojjzNkRsZWJiY1VGc28>

SOLUTION :

import org.apache.log4j.{Level, Logger}

import org.apache.spark.sql.SparkSession

object Assignment\_22\_2 extends App {

Logger.getLogger("org").setLevel(Level.OFF)

Logger.getLogger("akka").setLevel(Level.OFF)

val spark = SparkSession.builder

.master("local")

.appName("example")

.config("spark.sql.warehouse.dir","C://ACADGILD")

.getOrCreate()

// A CSV dataset is pointed to by path.

// The path can be either a single text file or a directory storing text files.

import spark.implicits.\_

val sc = spark.sparkContext

val tweets = sc.textFile("C:/ACADGILD/Big Data/demonetization-tweets.csv").map(\_.split(",")).filter(x=>x.length>=2)

.map(x => (x(0).replace("\"",""),x(1).replaceAll("\"","").toLowerCase))

.map(x => (x.\_1,x.\_2.split(" "))).toDF("id","words")

tweets.show()

tweets.createOrReplaceTempView("tweets")

val explode = spark.sql("select id as id,explode(words) as word from tweets").createOrReplaceTempView("tweet\_word")

val afinn = sc.textFile("C:/ACADGILD/Big Data/AFINN.txt").map(x => x.split("\t"))

.map(x => (x(0),x(1))).toDF("word","rating").createOrReplaceTempView("afinn")

val join = spark.sql("select t.id,AVG(a.rating) as rating from tweet\_word t join afinn a " +

"on t.word=a.word group by t.id order by rating desc").show}

OUTPUT:



