Report For Lab Assignment 5&6

1.

Question:

Spark and Smartphone/Watch Application
Implement a smart application with big data analytics related to your project showing the
collaboration between Spark and Smart Apps. Implement Twitter Streaming and perform
word count on it and publish the results and showcase it in your Smart Phone/Watch Application.

Description:

I have collected the famous top tweets by using twitter streaming and then performed wordcount on that set of tweets. After Performing wordcount I used Android Socket program to visualze that wordcount in the Android device. The following Screenshots explains all.

Screenshots:

```
// Print popular words
topCounts3.foreachRDD(rdd => {
  val topList = rdd.take(20)
  println("\nPopular words used in last 6 seconds (%s total):".format(rdd.count()))
  topList.foreach{case (count, word) => println("%s (%s times)".format(word, count))}


  var s:String="Popular words used in last 6 seconds (%s total): \nWords:Count \n"
  topList.foreach{case(count,word)=>{

    s+=word+" : "+count+"\n"

  }}
  SocketClient.sendCommandToRobot(s)
})
```

```
val sparkConf = new SparkConf().setAppName("SparkWordCount").setMaster("local[*]")

val sc=new SparkContext(sparkConf)

val input=sc.textFile("input")

val wc=input.flatMap(line=>{line.split(" ")}).map(word=>(word,1)).cache()

val output=wc.reduceByKey(_+_)
```

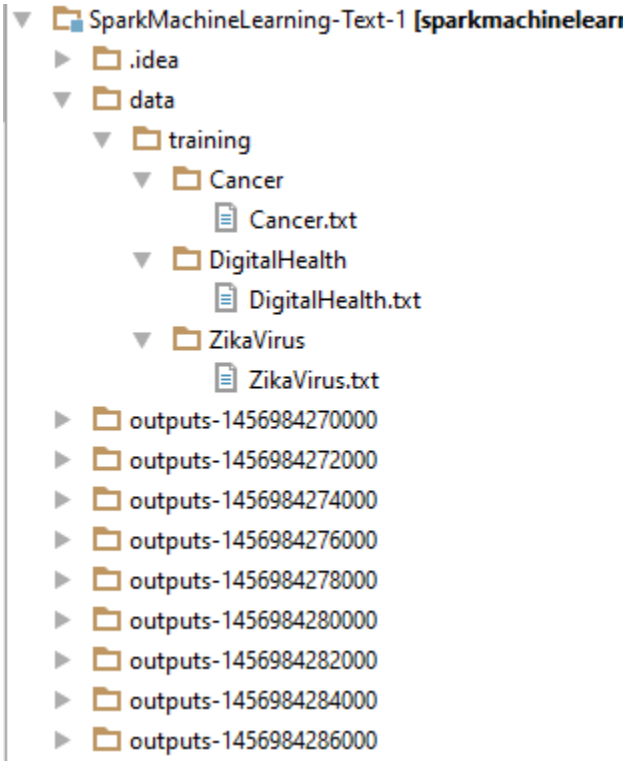
2.

Question:

Spark ML Lib Application
Perform a machine learning algorithm with the Twitter Streaming data to categorize each Tweet
1) Training datasets: Collect different categories of Tweets related to your project. (Categories can be based on HashTags / Subjects etc.)
2) Test data: the upcoming twitter stream

Description:

Firstly I have collected tweets based on the Hash tags in the json format using the twitter4j and the sample DotNet code as part of my training data, which are included in the Submission report.After collecting the data I have loaded the data in the following manner.



And then Collected the training data using spark live streaming

```
System.setProperty("twitter4j.oauth.consumerKey", "amWG9MI44iUDIC5traWtn3LYw")
System.setProperty("twitter4j.oauth.consumerSecret", "awSKM2jTcUP6OxZ3zOcQ701sooTOYvyYnNpvFF6Mm14qbqiqap")
System.setProperty("twitter4j.oauth.accessToken", "476949496-6bwiduPX98YEXZ6IhRNTQdGCnaDxsi9lLnxsIM3f")
System.setProperty("twitter4j.oauth.accessTokenSecret", "wME1q7RRlFiXE6u4ZswJW8HL8SUxIJKQPgCAH3uGicmxJ")

val stream = TwitterUtils.createStream(ssc, None, filters)

 stream.saveAsTextFiles("outputs")
```

Compared the Training data and Test data and Predict the Testing data Using Naive Bayes.

Screenshots:

```
    val lines=stream.map(status => status.getText())
    val data = lines.map(line => {

        val test = createLabeledDocumentTest(line, labelToNumeric, stopWords)
        test
    }
    )
  data.foreachRDD(rdd => {val X_test = tfidfTransformerTest(sc, rdd)
  val predictionAndLabel = model.predict(X_test)
  println("PREDICTION")
  predictionAndLabel.foreach(x => {
    labelToNumeric.foreach { y => if (y._2 == x) {
      println(y._1)
    }
    }
  })})
     ssc.start()
     ssc.awaitTermination()
     //.............................
```

```
16/03/02 23:51:25 INFO TaskSetManager: Finished task 2.0 in stage 29.0 (TID 100) in 10 ms on localhost (2/5)
16/03/02 23:51:25 INFO TaskSetManager: Finished task 1.0 in stage 29.0 (TID 99) in 11 ms on localhost (3/5)
16/03/02 23:51:25 INFO Executor: Running task 4.0 in stage 29.0 (TID 102)
16/03/02 23:51:25 INFO BlockManager: Found block rdd_88_3 locally
16/03/02 23:51:25 INFO BlockManager: Found block rdd_88_3 locally
16/03/02 23:51:25 INFO BlockManager: Found block rdd_88_4 locally
16/03/02 23:51:25 INFO BlockManager: Found block rdd_88_4 locally
16/03/02 23:51:25 INFO Executor: Finished task 3.0 in stage 29.0 (TID 101). 2044 bytes result sent to driver
DigitalHealth
ZikaVirus
Cancer
DigitalHealth
DigitalHealth
DigitalHealth
DigitalHealth
Cancer
DigitalHealth
Cancer
16/03/02 23:51:25 INFO TaskSetManager: Finished task 3.0 in stage 29.0 (TID 101) in 9 ms on localhost (4/5)
ZikaVirus
DigitalHealth
Cancer
DigitalHealth
DigitalHealth
Cancer
DigitalHealth
DigitalHealth
Cancer
```