

# **CS5542 : Lab Assignment #5&6**

Due on Wednesday, March 02, 2016

*Yugyung Lee*

**Sudhakar**

## 1: Spark and Smartphone/Watch Application

### Question:

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:

For this assignment, I have implemented the Twitter Streaming application by creating authentication details (Consumer secret, OAuth details) on <https://dev.twitter.com> and used word count program to modify the streaming application to count the words in the tweets which stream in the window of every 10sec. Used socket connection program to send the words and the corresponding count to the smartphone from apache spark.

## 2: Spark ML Lib Application Perform a machine learning

### Question:

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:

I have referred to the example for spark streaming application and modified the application to collect tweets for 10sec(Training data) in a series and placed it in the appropriate folders according to the keywords. I tried to make changes to the application to stream the incoming data and categorise the incoming data as per training data. But i could not complete the real time twitter streaming for prediction. So for the current time being I have implemented the logic to break the incoming twitter stream data into data blocks and feed them to the prediction model. In my observations I found that the matching results obtained from prediction were almost accurate.

### Reference:

I have went through the given examples in the class CS5542 by Dr. Yugyung Lee

**Screenshots:**

Q1:



Figure 1: TweetWordCount on smartphone

Q2:

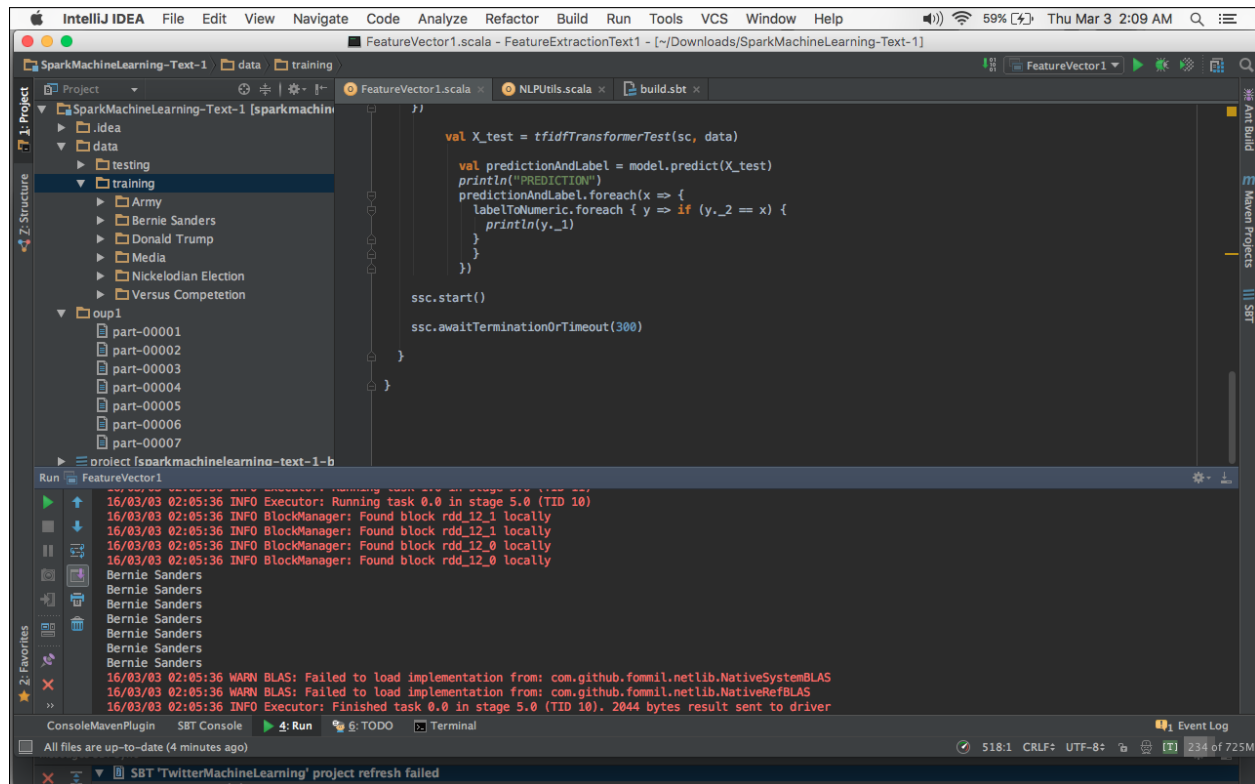


Figure 2: Machine learning algorithm implementation