# Lab 5 and 6 report

## Task 1

There were two tasks to be performed this time. The first task was to create an Android client and interact with Spark engine to perform word count on Twitter streaming data.
To achieve this task the following was done.

- Created an Android client.
- Implemented a Socket server at the client end.
- On press of the button the server would start and would be ready to listen to data.
- Created Twitter API key and secret.
- Utilized the Twitter Stream to get Tweets and filter the hashtags out of it.
- Then performed count of Tweets for 5 second window to get the the count of each hashtag in descending order.
- Send the top 5 count of hashtags to Android through Socket and display the result.

The screen shots for the task are given below.

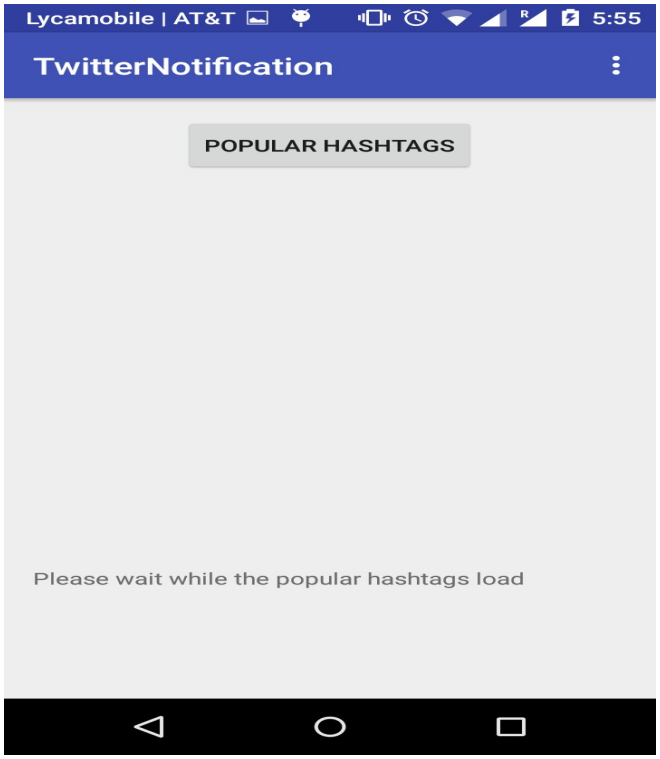


Fig 1 : Inital screen when the server starts

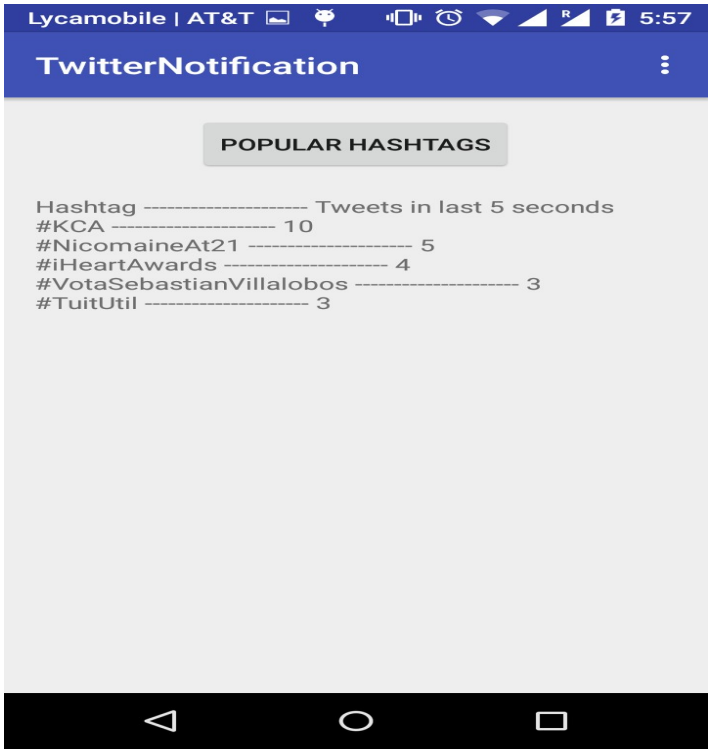


Fig 2 : The screen showing the top 5 trending hashtags with the count of Tweets.

# Task 2

This task was to create a classification system that can categorize Tweets to predefined classes.
In my case I had the training data which was set of Tweets on the following categories

1. #hillary
2. #uselections
3. #DonaldTrump

I collected a subset of live Tweets and allocated it as testing data.
Then I processed the testing to the FeatureVector algorithm to get it to predict the class to which the Tweet belongs to.
PFB the screenshot for the final output which classifies a particular set of Tweets to one class.