

Assignment 5 and 6

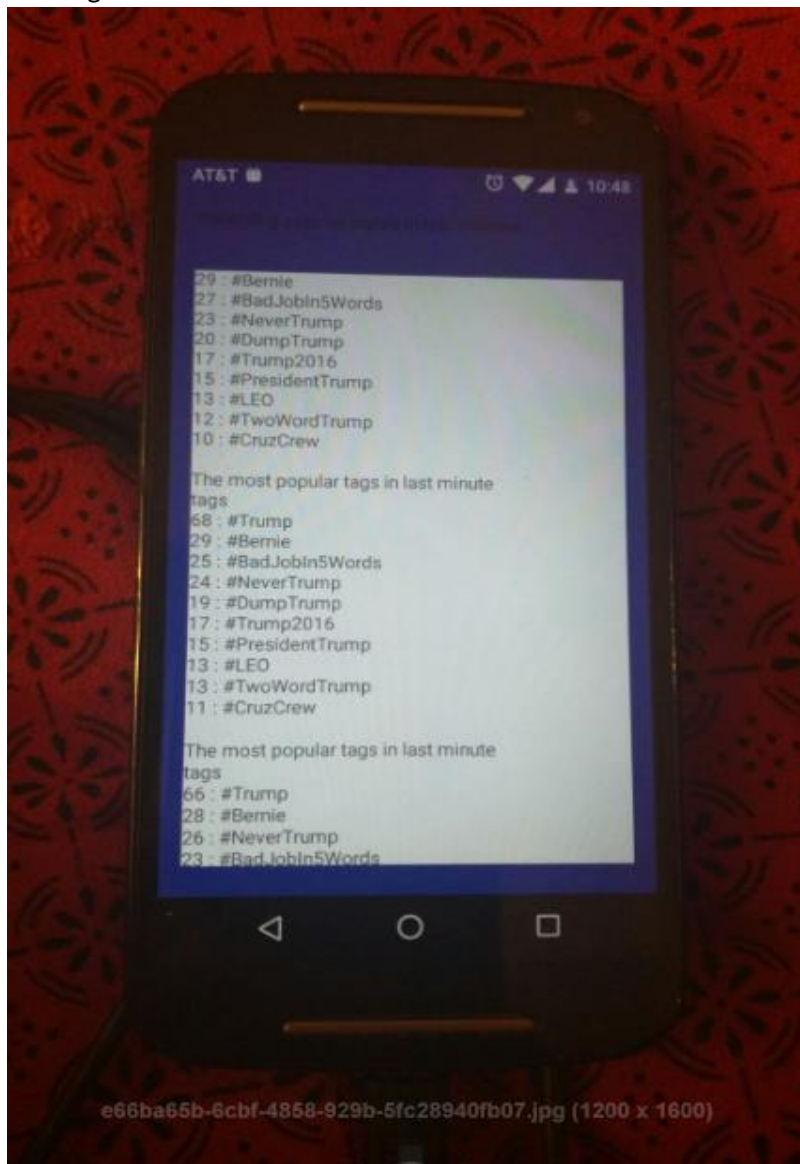
Q-1

I have collected the twitter tweet for by filtering it with following keywords as Oscar award and primary presidential election were there.

"oscar", "leo", "Primary", "election", "clinton", "cruz", "trump", "sanders"

I kept interval of 1 Minute and collecting the popular hashtags generated during last minute. I am sending this hash tag and its count result to the smart phone application.

Below is the output of how it looks to the mobile app screen. It keeps appending on the screen and when we scroll down we will get the latest tweets. First part is number of count and second part is hashtag.



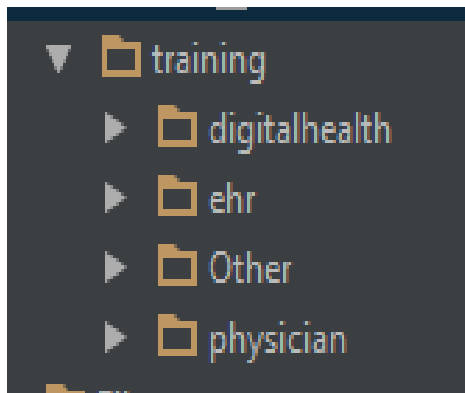
Q-2 : For this task I have divided it into 3 steps. First step is of collecting the dataset, second step is of examining tweet and train the model and the third and last step is to apply the model and predict the data. Below is the brief explanation.

We are trying to make recommendation system to recommend user some important tweets on basis of some famous medical hashtags. We have predefined our filter hashtags and filtering tweets in streaming. Below is the filter file.



allfilter.txt

For now we have divided it into below mentioned 4 categories. These categories are on basis of hashtags:



Below is the sample predicted results for few tweets.



61257.txt



61258.txt



61256.txt

Above three files are the collected tweets during streaming and below is the predicted output:

FeatureVector1.scala61258.txtMainClass.scalabuild.sbtUtils.scala

testingtest61256.txt61257.txt61258.txttrainingdigitalhealthehrother

708520417393319936::RT @MedDataInc: Scratching your head about @EveryICD10 codes? Learn the ABC's of #ICD10 <https://t.co/786eS6sEYR> <https://t.co/WeApply: https://t.co/dFK2y52Uei>
#OB-GYN - #CA #Modesto #PhysicianJobs #hiring #JobOpening #rtjobs <https://t.co/5Yk66sl5Tx>::physician708520869836918785::RT @NWrightDesignCo: Graphic

#digitalhealth #fitness #StockyRT <https://t.co/IBrtAgeByi>::708522545046130690::#Physician - OB/GYN
Apply: <https://t.co/w58HgAGBK1>
#OB-GYN - #CA #Fresno #PhysicianJobs #hiring #JobOpening #rtjobs <https://t.co/xCi8pHReWB>::physician708522618136166401::Check out this #job: #Recruit
Apply: <https://t.co/BjTG9AdyOS>
- #Bakersfield #CA #PhysicianJobs #hiring #JobOpening #rtjobs <https://t.co/xEbFsm9mOF>::physician708523282232954883::RT @Pallimed: Want to develop

Run: MainClassFeatureVector1

16/03/11 23:12:56 INFO DAGScheduler: Parents of final stage: List()
16/03/11 23:12:56 INFO DAGScheduler: Missing parents: List()
16/03/11 23:12:56 INFO DAGScheduler: Submitting ResultStage 5 (MapPartitionsRDD[17] at mapPartitions at NaiveBayes.scala:90), which has no missing parents
16/03/11 23:12:56 INFO MemoryStore: ensureFreeSpace(6144) called with curMem=88171159, maxMem=2050605711
16/03/11 23:12:56 INFO MemoryStore: Block broadcast_11 stored as values in memory (estimated size 6.0 KB, free 1871.5 MB)
16/03/11 23:12:56 INFO MemoryStore: ensureFreeSpace(3702) called with curMem=88177303, maxMem=2050605711
16/03/11 23:12:56 INFO MemoryStore: Block broadcast_11_piece0 stored as bytes in memory (estimated size 3.6 KB, free 1871.5 MB)
16/03/11 23:12:56 INFO BlockManagerInfo: Added broadcast_11_piece0 in memory on localhost:53115 (size: 3.6 KB, free: 1951.8 MB)
16/03/11 23:12:56 INFO SparkContext: Created broadcast 11 from broadcast at DAGScheduler.scala:861
16/03/11 23:12:56 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 5 (MapPartitionsRDD[17] at mapPartitions at NaiveBayes.scala:90)
16/03/11 23:12:56 INFO TaskSchedulerImpl: Adding task set 5.0 with 1 tasks
16/03/11 23:12:56 INFO TaskSetManager: Starting task 0.0 in stage 5.0 (TID 9, localhost, PROCESS_LOCAL, 2729 bytes)
16/03/11 23:12:56 INFO Executor: Running task 0.0 in stage 5.0 (TID 9)
16/03/11 23:12:56 INFO BlockManager: Found block rdd_12_0 locally
16/03/11 23:12:56 INFO BlockManager: Found block rdd_12_0 locally
16/03/11 23:12:56 INFO Executor: Finished task 0.0 in stage 5.0 (TID 9). 2044 bytes result sent to driver
physician
Other
Other
16/03/11 23:12:56 INFO TaskSetManager: Finished task 0.0 in stage 5.0 (TID 9) in 10 ms on localhost (1/1)