Technology :   Python: 2.7   Scala: 2.11   Spark: 2.2.1

Directory:

Surbhi_Batra_hw5
- ● Surbhi_Batra
  - ○ Surbhi_Batra_Description
- ● OutputFiles
- ● Solution
  - ○ Surbhi_Batra_DGIMAlgorithm.py
  - ○ Surbhi_Batra_TwitterStreaming.py
  - ○ Surbhi_Batra_DGIMAlgorithm.scala
  - ○ Surbhi_Batra_TwitterStreaming.scala
  - ○ Surbhi_Batra_hw5.jar

Algorithm:

>>>Execute these Commands after : cd/to/Surbhi_Batra_hw5/Solution

Python:

TASK : 1

TWITTER STREAMING
<pip install tweepy>

PYTHON:

Command :

$SPARK_HOME/bin/spark-submit Surbhi_Batra_TwitterStreaming.py

SCALA:

spark-submit --class TwitterStreaming
/home/surbhi/IdeaProjects/SB_hw5/out/artifacts/SB_hw5_jar/Surbhi_Batra_hw5.jar

```
Number of twitter from the beginning: 111
Top 5 hot hashtags:
killer:1
ALDUBxEBPamana:1
빅스:1
이재환:1
민규:1
The average length of the twitter is: 132.8018


Number of twitter from the beginning: 120
Top 5 hot hashtags:
killer:1
ALDUBxEBPamana:1
빅스:1
이재환:1
민규:1
The average length of the twitter is: 124.8
```

TASK : 2
DGIM ALGORITHM:
PYTHON:

Command:

1) To run Simulation:
 java −cp <Path of Streaming Simulation.jar> Streaming Simulation 9999 10
java -cp
Documents/2018Summer/DM/Assignment_05/Description/Data/streaming4inf553assignment5.j
ar StreamingSimulation 9999 100

2) To run the file:
$SPARK_HOME/bin/spark-submit Surbhi_Batra_DGIMAlgorithm.py

SCALA:

spark-submit --class DGIMAlgorithm
/home/surbhi/IdeaProjects/SB_hw5/out/artifacts/SB_hw5_jar/Surbhi_Batra_hw5.jar

```
Estimated number of ones in the last 1000 bits:353
Actual number of ones in the last 1000 bits: 417

Estimated number of ones in the last 1000 bits:401
Actual number of ones in the last 1000 bits: 465
```