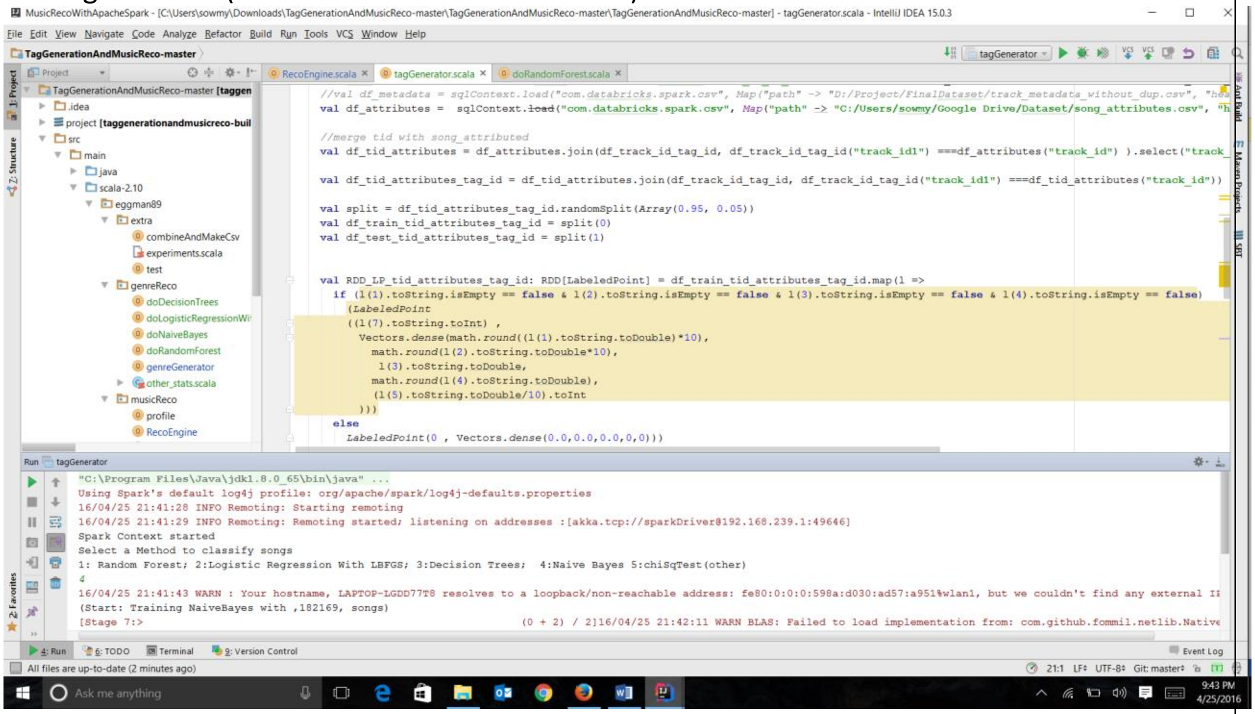
Varun Chavakula(4) Sowmya yelmati(32)

Lab 10 Assignment

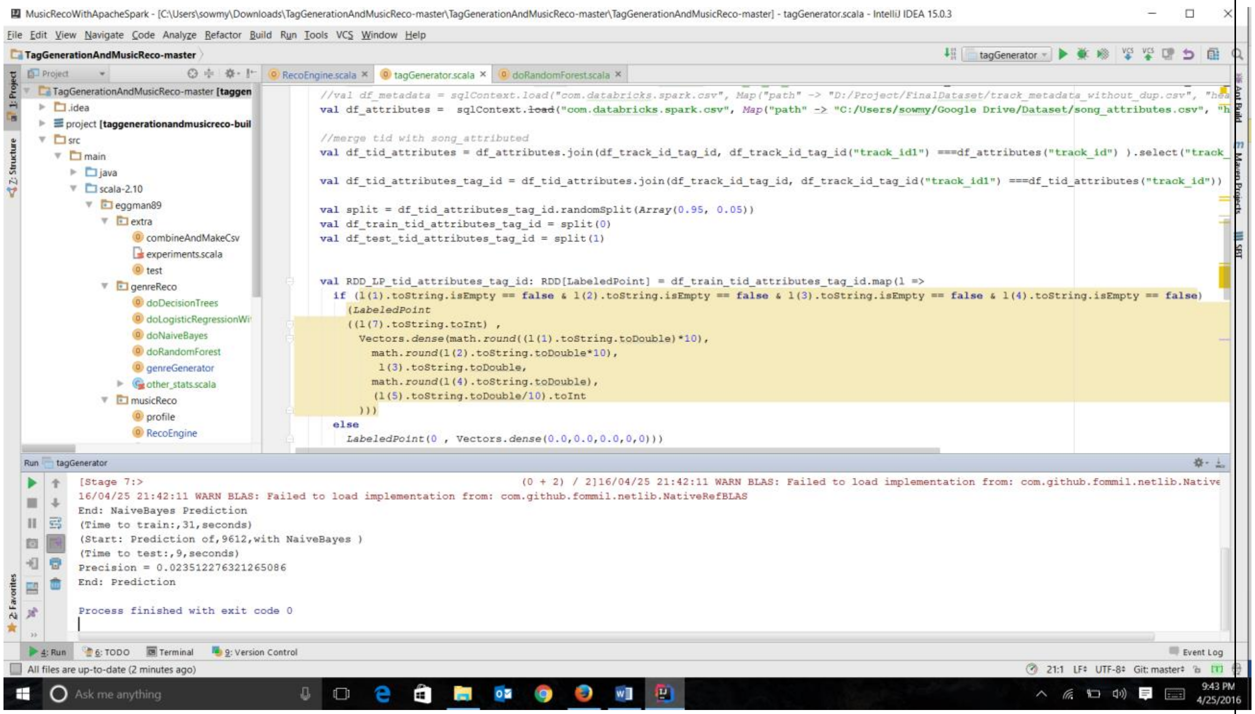
Audio Classification and Smart Phone Notification

1) Audio collection based on the tags (related to your project)
   a. Training Datasets (audio data from smart devices or online sound repositories such as https://www.freesound.org/)
   b. Testing Datasets (audio data from smart devices)


MusicRecoWithApacheSpark - [C:\Users\sowmy\Downloads\TagGenerationAndMusicReco-master\TagGenerationAndMusicReco-master\TagGenerationAndMusicReco-master] - tagGenerator.scala - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
TagGenerationAndMusicReco-master
tagGenerator
RecoEngine.scala
tagGenerator.scala
doRandomForest.scala
Project
TagGenerationAndMusicReco-master [taggen
.idea
project [taggenerationandmusicreco-buil
src
main
java
scala-2.10
eggman89
extra
combineAndMakeCsv
experiments.scala
test
genreReco
doDecisionTrees
doLogisticRegressionWi
doNaiveBayes
doRandomForest
genreGenerator
other_stats.scala
musicReco
profile
RecoEngine
Run tagGenerator
4: Run
6: TODO
Terminal
9: Version Control
Event Log
All files are up-to-date (2 minutes ago)
21:1 LF UTF-8 Git: master
Ask me anything
9:43 PM
4/25/2016

```
//val df_metadata = sqlContext.load("com.databricks.spark.csv", Map("path" -> "D:/Project/FinalDataset/track_metadata_without_dup.csv", "hea
val df_attributes =  sqlContext.~~load~~("com.databricks.spark.csv", Map("path" -> "C:/Users/sowmy/Google Drive/Dataset/song_attributes.csv", "h

//merge tid with song_attributed
val df_tid_attributes = df_attributes.join(df_track_id_tag_id, df_track_id_tag_id("track_id1") ===df_attributes("track_id") ).select("track_

val df_tid_attributes_tag_id = df_tid_attributes.join(df_track_id_tag_id, df_track_id_tag_id("track_id1") ===df_tid_attributes("track_id"))

val split = df_tid_attributes_tag_id.randomSplit(Array(0.95, 0.05))
val df_train_tid_attributes_tag_id = split(0)
val df_test_tid_attributes_tag_id = split(1)


val RDD_LP_tid_attributes_tag_id: RDD[LabeledPoint] = df_train_tid_attributes_tag_id.map(l =>
  if (l(1).toString.isEmpty == false & l(2).toString.isEmpty == false & l(3).toString.isEmpty == false & l(4).toString.isEmpty == false)
    (LabeledPoint
    ((l(7).toString.toInt) ,
      Vectors.dense(math.round((l(1).toString.toDouble)*10),
        math.round(l(2).toString.toDouble*10),
         l(3).toString.toDouble,
        math.round(l(4).toString.toDouble),
         (l(5).toString.toDouble/10).toInt
      )))
  else
    LabeledPoint(0 , Vectors.dense(0.0,0.0,0.0,0,0)))
```

```
"C:\Program Files\Java\jdk1.8.0_65\bin\java" ...
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
16/04/25 21:41:28 INFO Remoting: Starting remoting
16/04/25 21:41:29 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriver@192.168.239.1:49646]
Spark Context started
Select a Method to classify songs
1: Random Forest; 2:Logistic Regression With LBFGS; 3:Decision Trees;  4:Naive Bayes 5:chiSqTest(other)
4
16/04/25 21:41:43 WARN : Your hostname, LAPTOP-LGDD77T8 resolves to a loopback/non-reachable address: fe80:0:0:0:598a:d030:ad57:a951%wlan1, but we couldn't find any external IP
(Start: Training NaiveBayes with ,182169, songs)
[Stage 7:>                                                          (0 + 2) / 2]16/04/25 21:42:11 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.Native
```

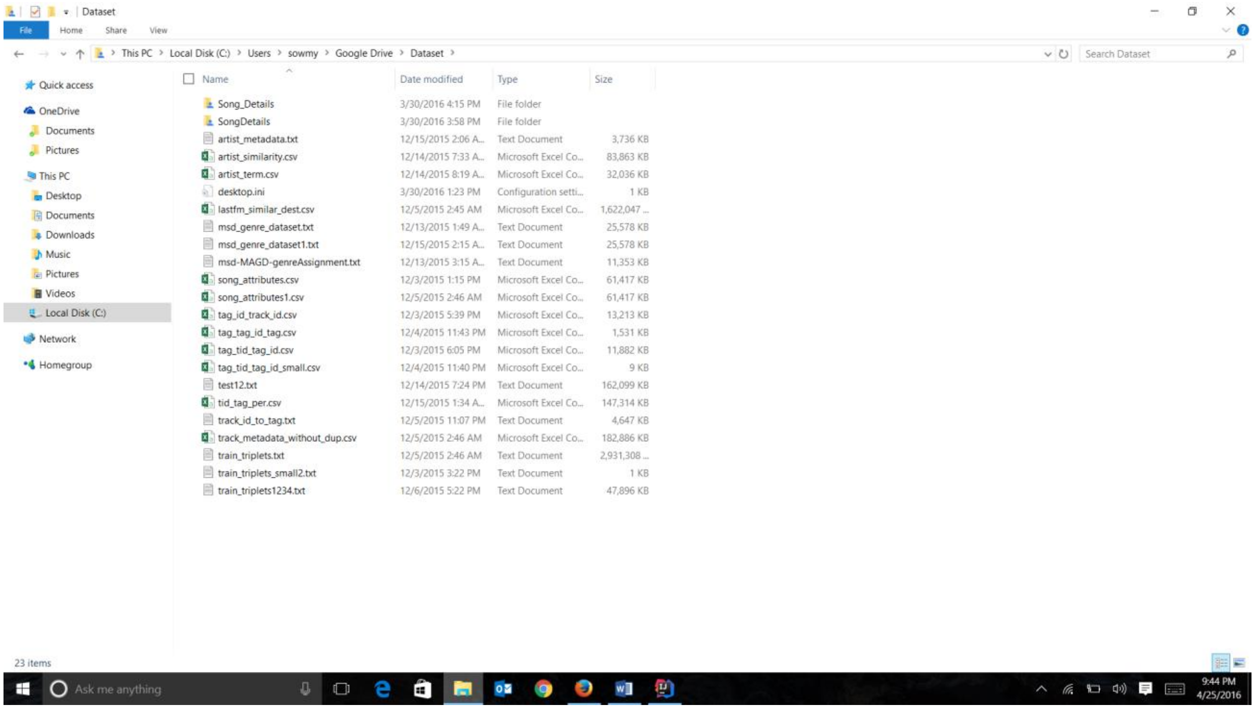


MusicRecoWithApacheSpark - [C:\Users\sowmy\Downloads\TagGenerationAndMusicReco-master\TagGenerationAndMusicReco-master\TagGenerationAndMusicReco-master] - tagGenerator.scala - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
TagGenerationAndMusicReco-master
tagGenerator
RecoEngine.scala
tagGenerator.scala
doRandomForest.scala
Project
TagGenerationAndMusicReco-master [taggen
.idea
project [taggenerationandmusicreco-buil
src
main
java
scala-2.10
eggman89
extra
combineAndMakeCsv
experiments.scala
test
genreReco
doDecisionTrees
doLogisticRegressionWi
doNaiveBayes
doRandomForest
genreGenerator
other_stats.scala
musicReco
profile
RecoEngine
Run tagGenerator
4: Run
6: TODO
Terminal
9: Version Control
Event Log
All files are up-to-date (2 minutes ago)
21:1 LF‡ UTF-8‡ Git: master‡
Ask me anything
9:43 PM
4/25/2016

```
//val df_metadata = sqlContext.load("com.databricks.spark.csv", Map("path" -> "D:/Project/FinalDataset/track_metadata_without_dup.csv", "hea
val df_attributes =  sqlContext.~~load~~("com.databricks.spark.csv", Map("path" -> "C:/Users/sowmy/Google Drive/Dataset/song_attributes.csv", "h

//merge tid with song_attributed
val df_tid_attributes = df_attributes.join(df_track_id_tag_id, df_track_id_tag_id("track_id1") ===df_attributes("track_id") ).select("track_

val df_tid_attributes_tag_id = df_tid_attributes.join(df_track_id_tag_id, df_track_id_tag_id("track_id1") ===df_tid_attributes("track_id"))

val split = df_tid_attributes_tag_id.randomSplit(Array(0.95, 0.05))
val df_train_tid_attributes_tag_id = split(0)
val df_test_tid_attributes_tag_id = split(1)


val RDD_LP_tid_attributes_tag_id: RDD[LabeledPoint] = df_train_tid_attributes_tag_id.map(l =>
  if (l(1).toString.isEmpty == false & l(2).toString.isEmpty == false & l(3).toString.isEmpty == false & l(4).toString.isEmpty == false)
    (LabeledPoint
    ((l(7).toString.toInt) ,
      Vectors.dense(math.round((l(1).toString.toDouble)*10),
        math.round(l(2).toString.toDouble*10),
         l(3).toString.toDouble,
        math.round(l(4).toString.toDouble),
         (l(5).toString.toDouble/10).toInt
      )))
  else
    LabeledPoint(0 , Vectors.dense(0.0,0.0,0.0,0,0)))
```

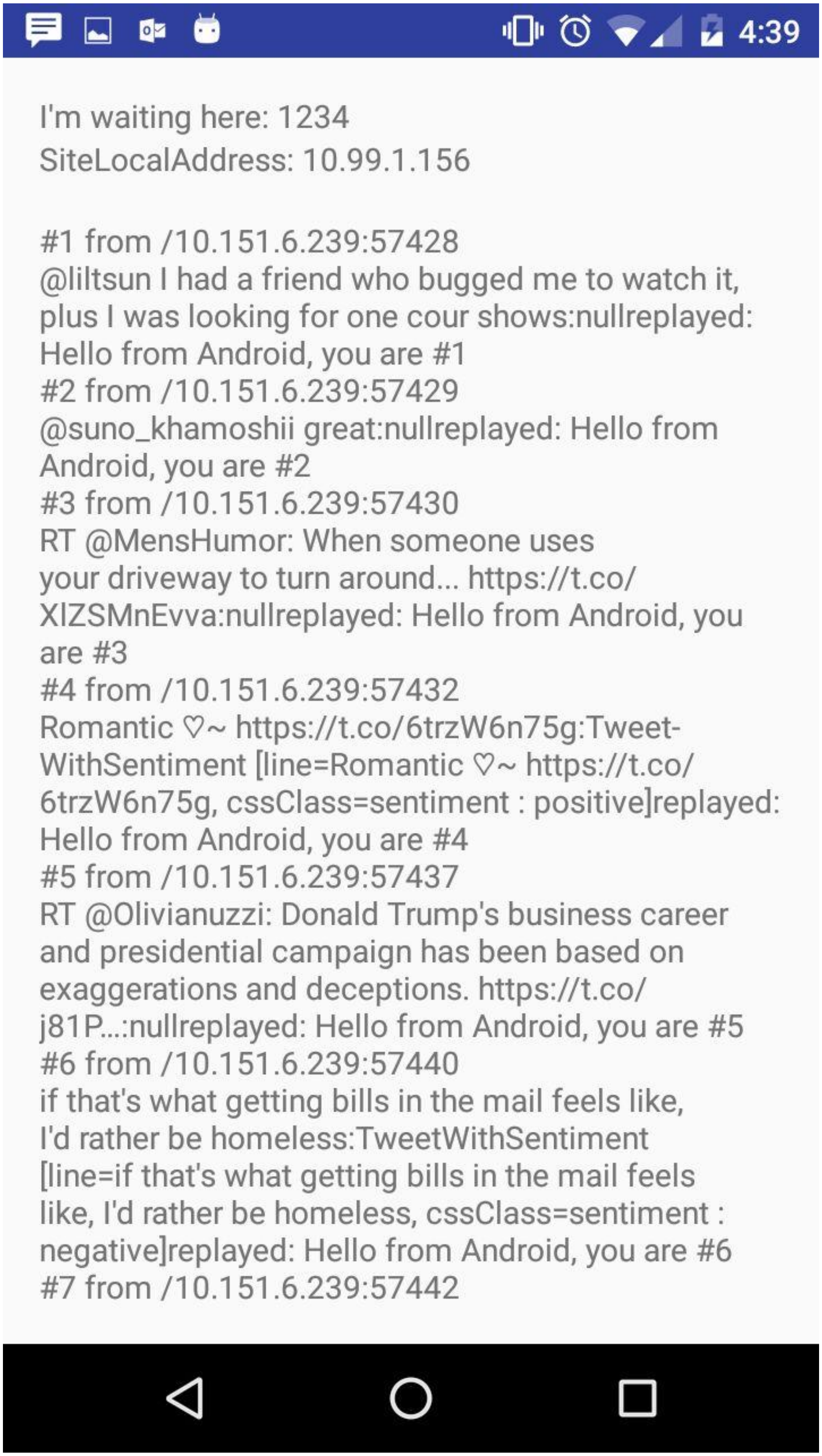
```
[Stage 7:>                                                          (0 + 2) / 2]16/04/25 21:42:11 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.Native
16/04/25 21:42:11 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
End: NaiveBayes Prediction
(Time to train:,31,seconds)
(Start: Prediction of,9612,with NaiveBayes )
(Time to test:,9,seconds)
Precision = 0.023512276321265086
End: Prediction

Process finished with exit code 0
```

2) Audio Classification based on the categories related to your project



Name	Date modified	Type	Size
Song_Details	3/30/2016 4:15 PM	File folder	
SongDetails	3/30/2016 3:58 PM	File folder	
artist_metadata.txt	12/15/2015 2:06 A...	Text Document	3,736 KB
artist_similarity.csv	12/14/2015 7:33 A...	Microsoft Excel Co...	83,863 KB
artist_term.csv	12/14/2015 8:19 A...	Microsoft Excel Co...	32,036 KB
desktop.ini	3/30/2016 1:23 PM	Configuration setti...	1 KB
lastfm_similar_dest.csv	12/5/2015 2:45 AM	Microsoft Excel Co...	1,622,047 ...
msd_genre_dataset.txt	12/13/2015 1:49 A...	Text Document	25,578 KB
msd_genre_dataset1.txt	12/15/2015 2:15 A...	Text Document	25,578 KB
msd-MAGD-genreAssignment.txt	12/13/2015 3:15 A...	Text Document	11,353 KB
song_attributes.csv	12/3/2015 1:15 PM	Microsoft Excel Co...	61,417 KB
song_attributes1.csv	12/5/2015 2:46 AM	Microsoft Excel Co...	61,417 KB
tag_id_track_id.csv	12/3/2015 5:39 PM	Microsoft Excel Co...	13,213 KB
tag_tag_id_tag.csv	12/4/2015 11:43 PM	Microsoft Excel Co...	1,531 KB
tag_tid_tag_id.csv	12/3/2015 6:05 PM	Microsoft Excel Co...	11,882 KB
tag_tid_tag_id_small.csv	12/4/2015 11:40 PM	Microsoft Excel Co...	9 KB
test12.txt	12/14/2015 7:24 PM	Text Document	162,099 KB
tid_tag_per.csv	12/15/2015 1:34 A...	Microsoft Excel Co...	147,314 KB
track_id_to_tag.txt	12/5/2015 11:07 PM	Text Document	4,647 KB
track_metadata_without_dup.csv	12/5/2015 2:46 AM	Microsoft Excel Co...	182,886 KB
train_triplets.txt	12/5/2015 2:46 AM	Text Document	2,931,308 ...
train_triplets_small2.txt	12/3/2015 3:22 PM	Text Document	1 KB
train_triplets1234.txt	12/6/2015 5:22 PM	Text Document	47,896 KB

3) Notification to smartphone/smartwatch

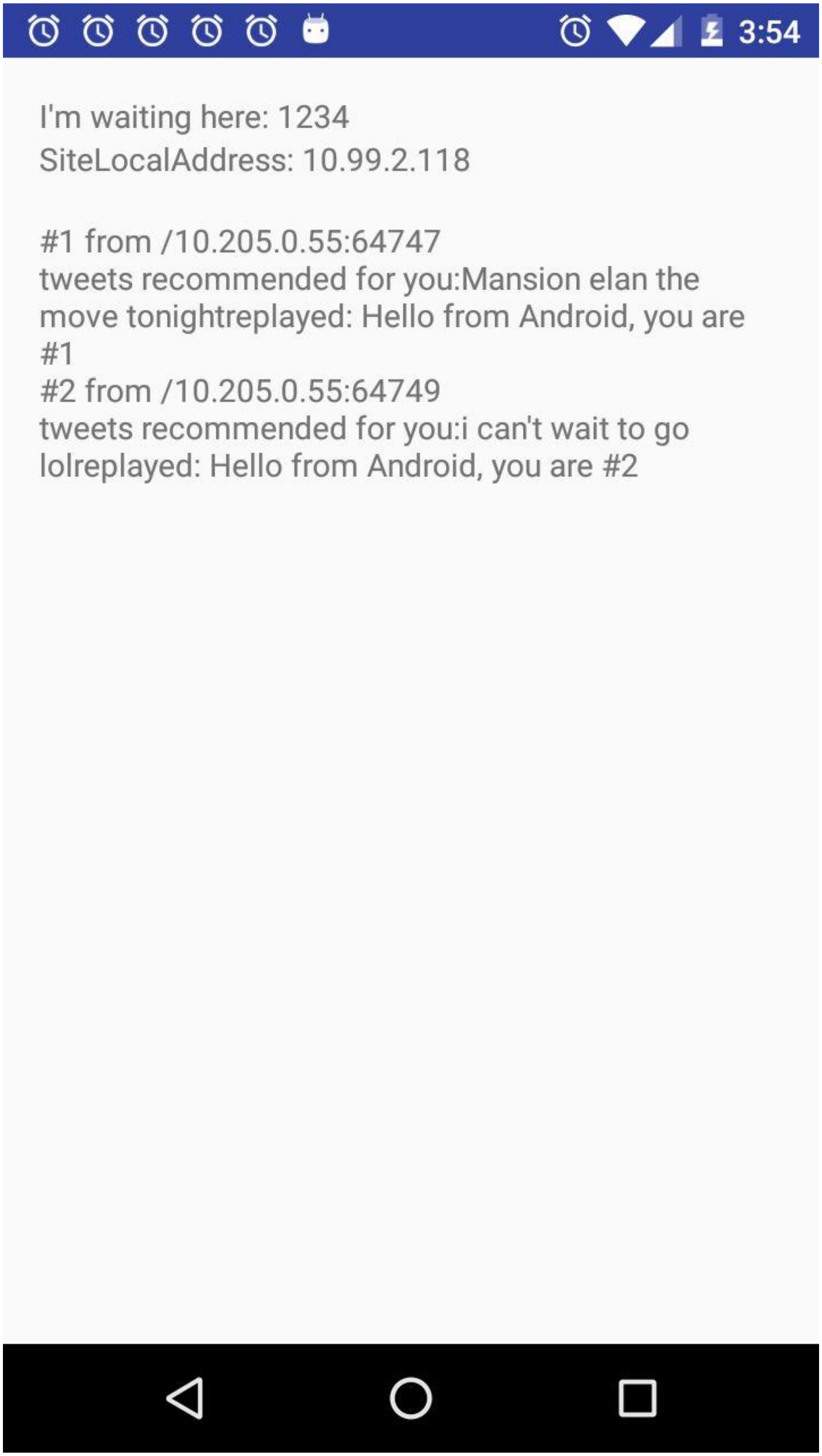


```
I'm waiting here: 1234
SiteLocalAddress: 10.99.1.156

#1 from /10.151.6.239:57428
@liltsun I had a friend who bugged me to watch it,
plus I was looking for one cour shows:nullreplayed:
Hello from Android, you are #1
#2 from /10.151.6.239:57429
@suno_khamoshii great:nullreplayed: Hello from
Android, you are #2
#3 from /10.151.6.239:57430
RT @MensHumor: When someone uses
your driveway to turn around... https://t.co/
XlZSMnEvva:nullreplayed: Hello from Android, you
are #3
#4 from /10.151.6.239:57432
Romantic ♡~ https://t.co/6trzW6n75g:Tweet-
WithSentiment [line=Romantic ♡~ https://t.co/
6trzW6n75g, cssClass=sentiment : positive]replayed:
Hello from Android, you are #4
#5 from /10.151.6.239:57437
RT @Olivianuzzi: Donald Trump's business career
and presidential campaign has been based on
exaggerations and deceptions. https://t.co/
j81P...:nullreplayed: Hello from Android, you are #5
#6 from /10.151.6.239:57440
if that's what getting bills in the mail feels like,
I'd rather be homeless:TweetWithSentiment
[line=if that's what getting bills in the mail feels
like, I'd rather be homeless, cssClass=sentiment :
negative]replayed: Hello from Android, you are #6
#7 from /10.151.6.239:57442
```

4) Bonus points:

Music Recommendation system (related to your own project)

a. make a recommendation based on user profile (e.g., preferences, location, gender, age)
b. Recommendation through smartphone/smartwatch using your ML application



```
I'm waiting here: 1234
SiteLocalAddress: 10.99.2.118

#1 from /10.205.0.55:64747
tweets recommended for you:Mansion elan the
move tonightreplayed: Hello from Android, you are
#1
#2 from /10.205.0.55:64749
tweets recommended for you:i can't wait to go
lolreplayed: Hello from Android, you are #2
```