CS5542 BIG DATA ANALYTICS AND APPS

Increment -3 Report (04/06/2016)

Project Group -7

By

Abhiram Ampabathina (1)

Harshini Medikonda (14)

Hirenbhai Harshadbhai Shah(27)

Dinesh Reddy (19)

I.INTRODUCTION:

The heart rate of a person depends on age, gender, daily physical activity, mental stress and many other activities/conditions. Furthermore, there is no proper equipment that can keep a track of heart beat rate. We intend to do a system that can collect the person's daily heart rate activity, store it in a database, analyze the heart rate and the activity the person is performing. Moreover, the application can analyze the data and recommend mental or physical activities to be performed by the user to keep the heart rate optimal. It can also suggest the timings of the abnormal heart rate. All these would give a clear idea of the medical condition of the user and the better usage of it can help in a longer life.

II.PROJECT GOAL AND OBJECTIVES:

OVERALL GOAL:

The goal of the project is to build a system that can take care of the user's health. This heart rate system is an android application which he can view even through the smart watch. This application works with the heart rate sensor embedded in the smart watch. It can observe the patterns of the heart rate and determine the health condition. It recommends the user with the necessary physical and mental activity.

SPECIFIC OBJECTIVE:

The objectives that would be achieved are as follows:

- Collect the heart rate and step count of the user
- Store the heart rate in regular intervals
- Get the heart rate onto HDFS per day basis
- Analyze it using machine learning algorithms.
- Notifying the health conditions using smart watch and smart phone
- Recommend the activities to be done by the user.
- Have a medical record, convenient and cost efficient.

SPECIFIC FEATURES:

The specific features designed in the project are:

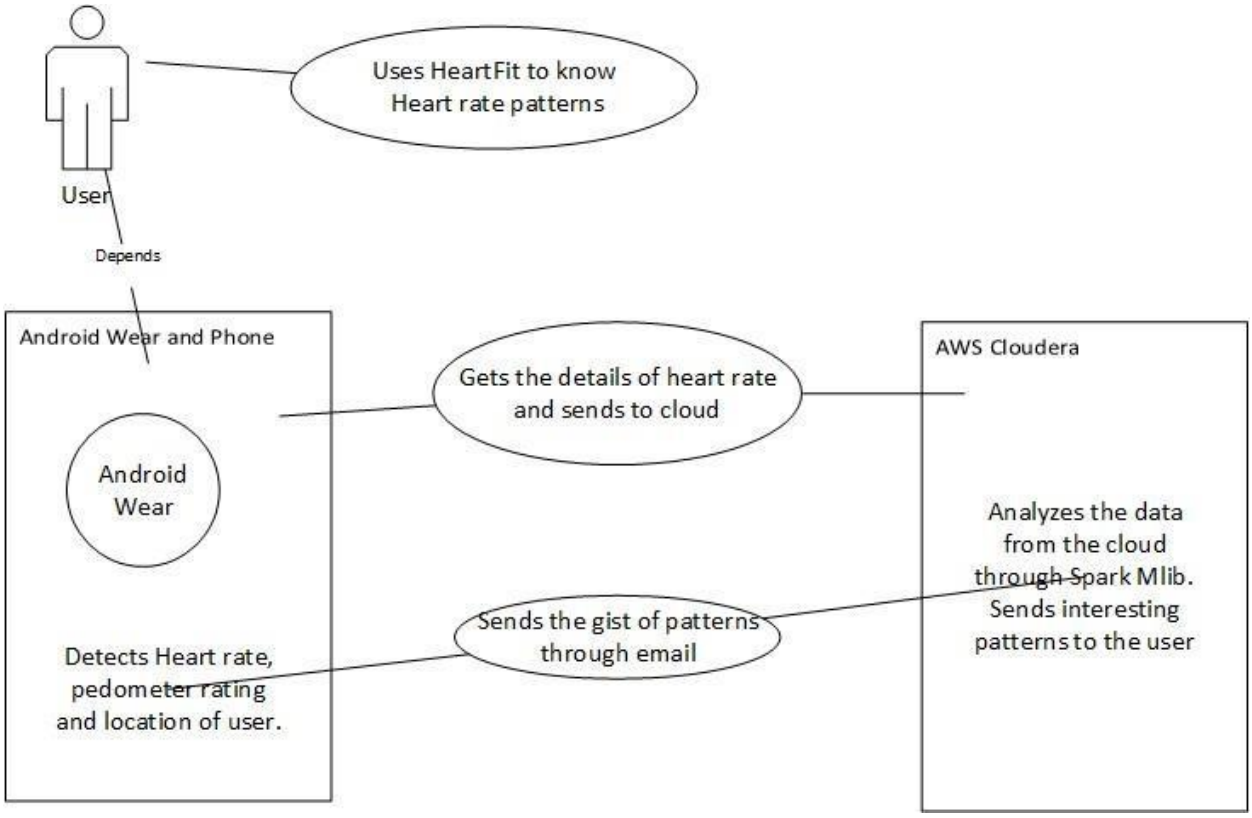
- Heart beat analysis
- Step count analysis
- Notification of current health condition
- Recommendation of health care

SIGNIFICANCE:

The main significance of this application is it is a system that is required for every person in their daily life. It is a trending smart application which makes the life easier. It is beneficial and becomes a part of the life in the upcoming years.

III. PROJECT PLAN:

1. Stories : Scenario & Use case specification



FEATURE DESIGN:

The application is designed to have the following features

1. Ability to run the application background all the time.

2. Ability to read heart rate accurately.

3. Ability to read the location of the user accurately.

4. Ability to send the information to cloud DB when internet is connected.

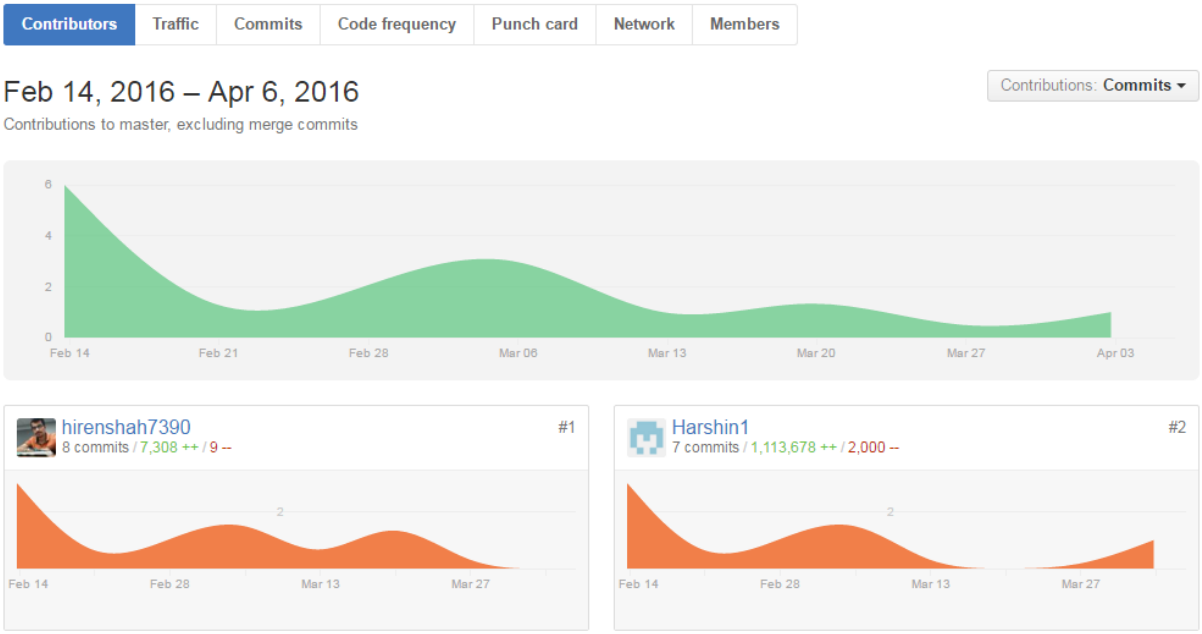
The system performs clustering based on the heart rate data of the user and recommends the user of the exercises to follow to stay healthy.

Each feature is designed accordingly to the specifications it should perform.

FEATURE IMPLEMENTATION:

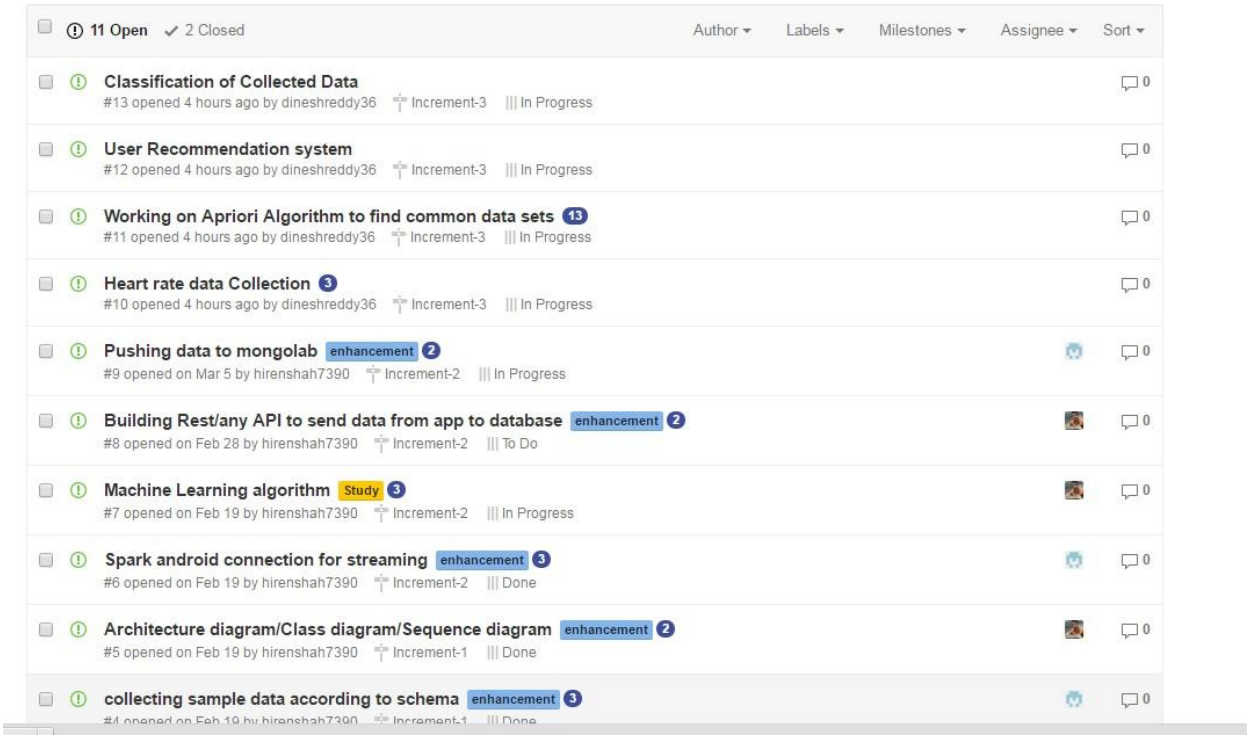
The application is designed to work continuously in the background. For this we added permission in android activity.xml . So, that the phone or the wear supports to run the application in background. To read user heart rate, we added heart rate sensor as a dependency to the application dependency list. This will accurately detect user's heart rate and return it to mobile. Our application will also send location details of user. For this we included GPS Location dependency plugin to android. This will send us the co-ordinates of the user's location. We also need to make sure the device is connected to internet in order to send the user's details to cloud DB, furtherly to analyze the data in cloudera.

2. PROJECT TIMELINES, MEMBERS, TASK RESOPNSIBILITY:



BURNDOWN CHARTS:

ISSUES:


11 Open
2 Closed
Author
Labels
Milestones
Assignee
Sort
Classification of Collected Data
#13 opened 4 hours ago by dineshreddy36 Increment-3 In Progress
0
User Recommendation system
#12 opened 4 hours ago by dineshreddy36 Increment-3 In Progress
0
Working on Apriori Algorithm to find common data sets 13
#11 opened 4 hours ago by dineshreddy36 Increment-3 In Progress
0
Heart rate data Collection 3
#10 opened 4 hours ago by dineshreddy36 Increment-3 In Progress
0
Pushing data to mongolab enhancement 2
#9 opened on Mar 5 by hirenshah7390 Increment-2 In Progress
0
Building Rest/any API to send data from app to database enhancement 2
#8 opened on Feb 28 by hirenshah7390 Increment-2 To Do
0
Machine Learning algorithm Study 3
#7 opened on Feb 19 by hirenshah7390 Increment-2 In Progress
0
Spark android connection for streaming enhancement 3
#6 opened on Feb 19 by hirenshah7390 Increment-2 Done
0
Architecture diagram/Class diagram/Sequence diagram enhancement 2
#5 opened on Feb 19 by hirenshah7390 Increment-1 Done
0
collecting sample data according to schema enhancement 3
0

IV. THIRD INCREMENT REPORT:

EXISTING API:

- MongoLab API:

https://api.mongolab.com/api/1/databases/my-db/collections?apiKey=myAPIKey

This API is used to store the heart rate and step count in the database and get the heart rate from the database.

- Heart Rate and Step counter Sensor

The heart rate and step counter sensors are embedded in the smart watch which can be used to get the data. This data is sent to the Spark HDFS system on a per-day basis.

- Java Mail API

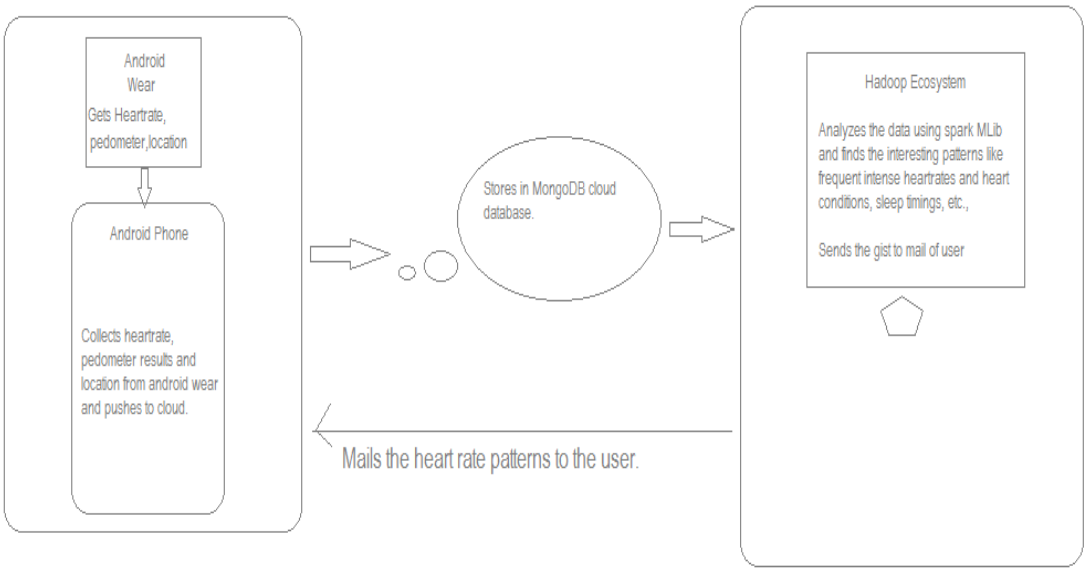
This is used to send an email of the results to the user as an email.

- Twitter Streaming API

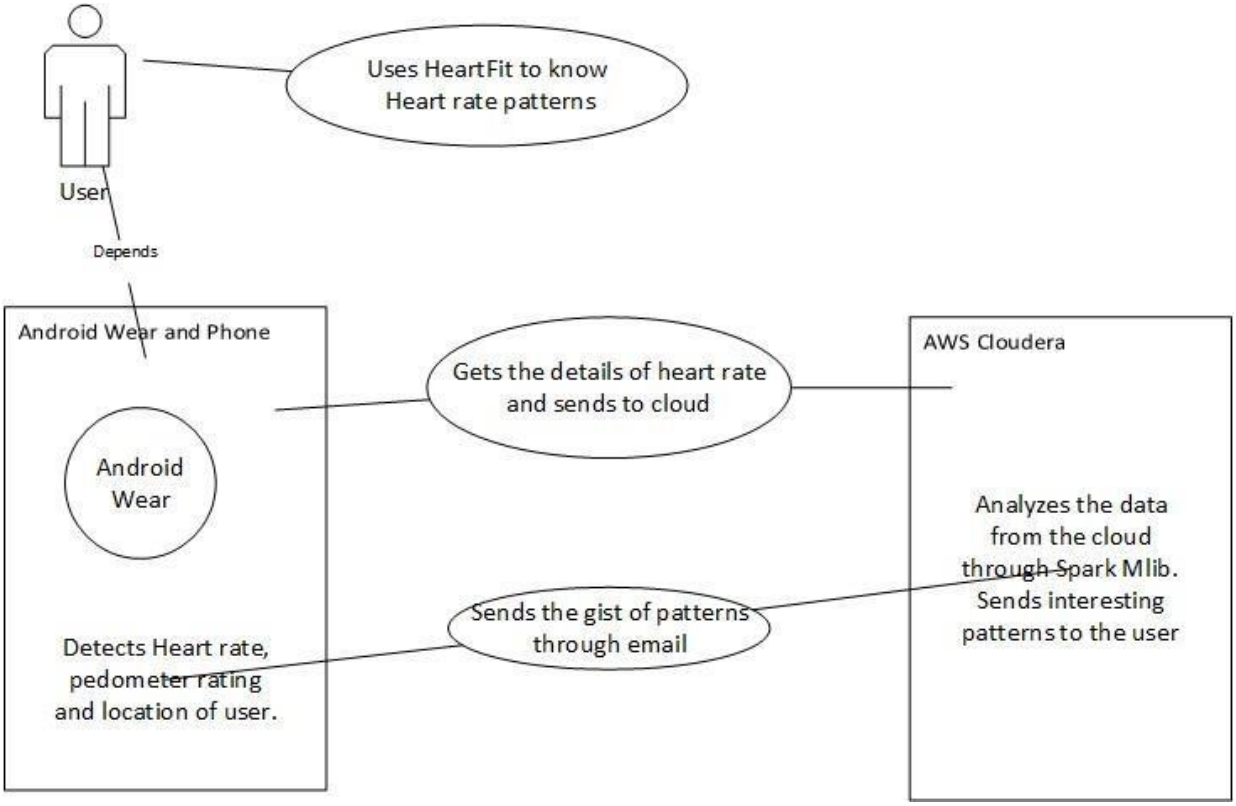
The tweets are collected as per the keywords heartrate, fitness, pulse, health and these are analyzed.

DESIGN OF FEATURES

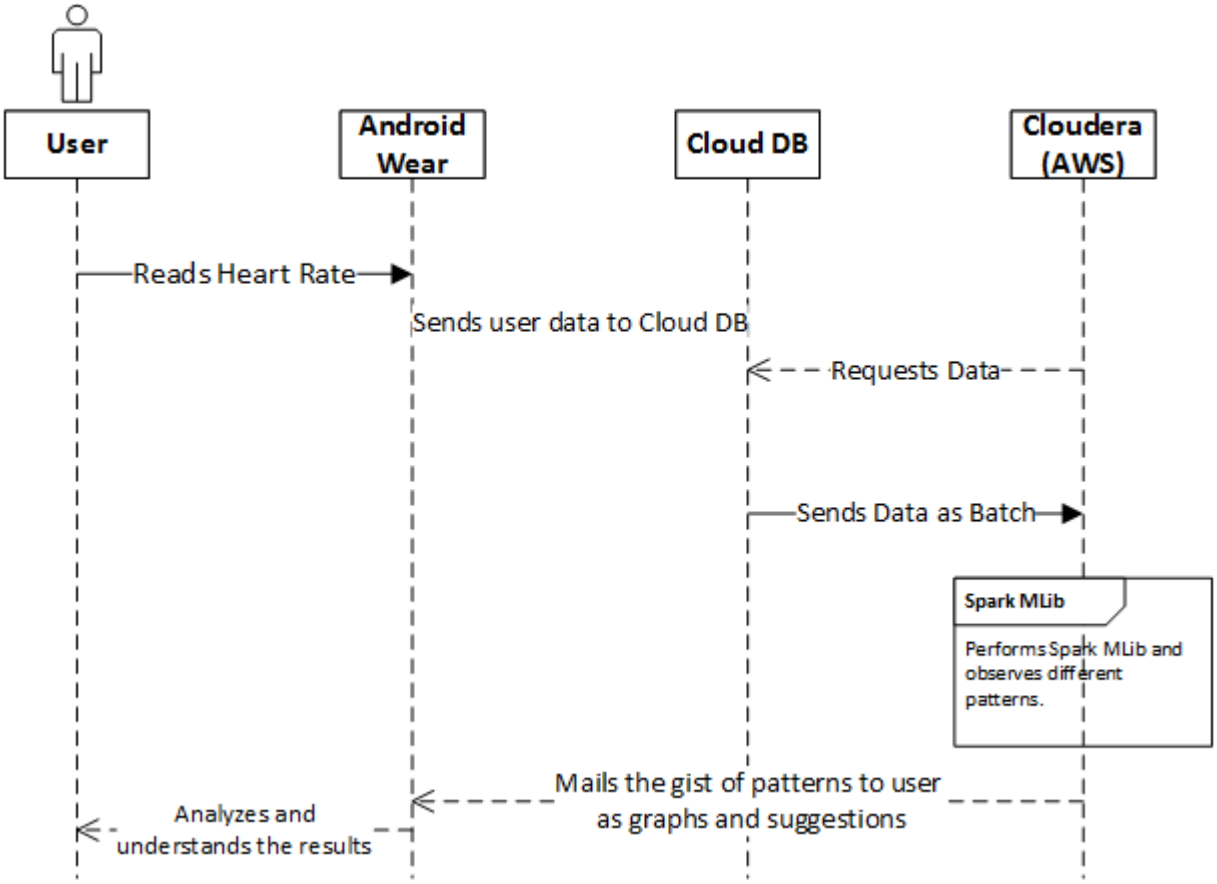
ARCHITECHTURE DIAGRAM:



CLASS DIAGRAM:



SEQUENCE DIAGRAM:

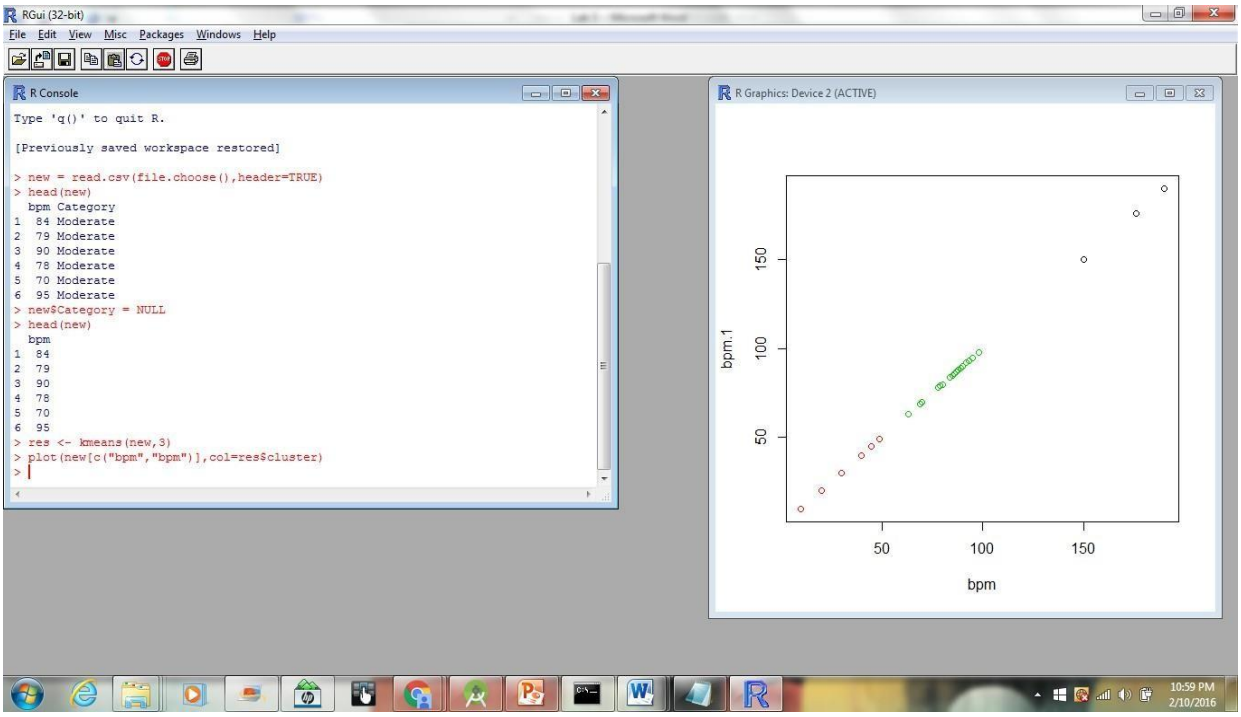


SPARK/MACHINE LEARNING ALGORITHMS:

In this application, we are planning to use the below machine learning algorithms to analyze the data.

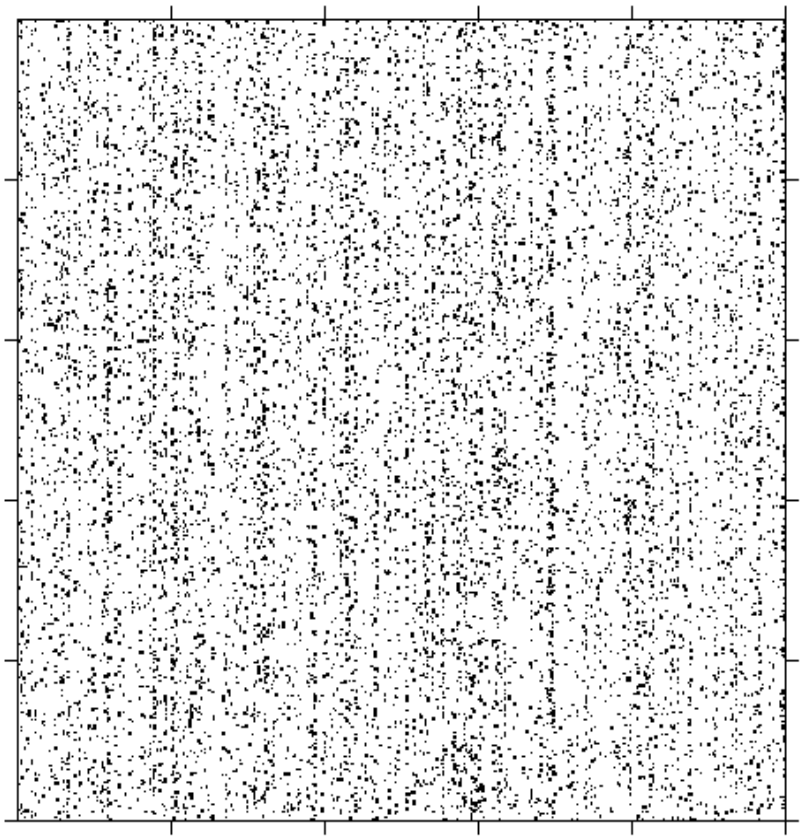
1. **K-means Clustering:**

K-means clustering algorithm is a cluster analysis in which k clusters are formed with n observations. The similar observations are clustered determining the centroid. Here the similar heart rates are clustered and the patterns are determined.



2. Apriori Algorithm:

The Apriori algorithm is an algorithm for mining frequent datasets. Since the heart rate of the user when collected it produces similar data and frequent datasets are formed. This enables the use of this algorithm to determine the patterns.



DATASETS:

The datasets in the Heartfit application consists of heart rate data, the steps walked for the day, the timing, the geolocation where it is captured. These datasets are analyzed with the machine learning algorithms and the corresponding patterns are generated.

It would appear as Step Count, Heart rate, time stamp, geolocation. Few more features can be added.

IMPLEMENTATION:

Mobile Client Implementation:

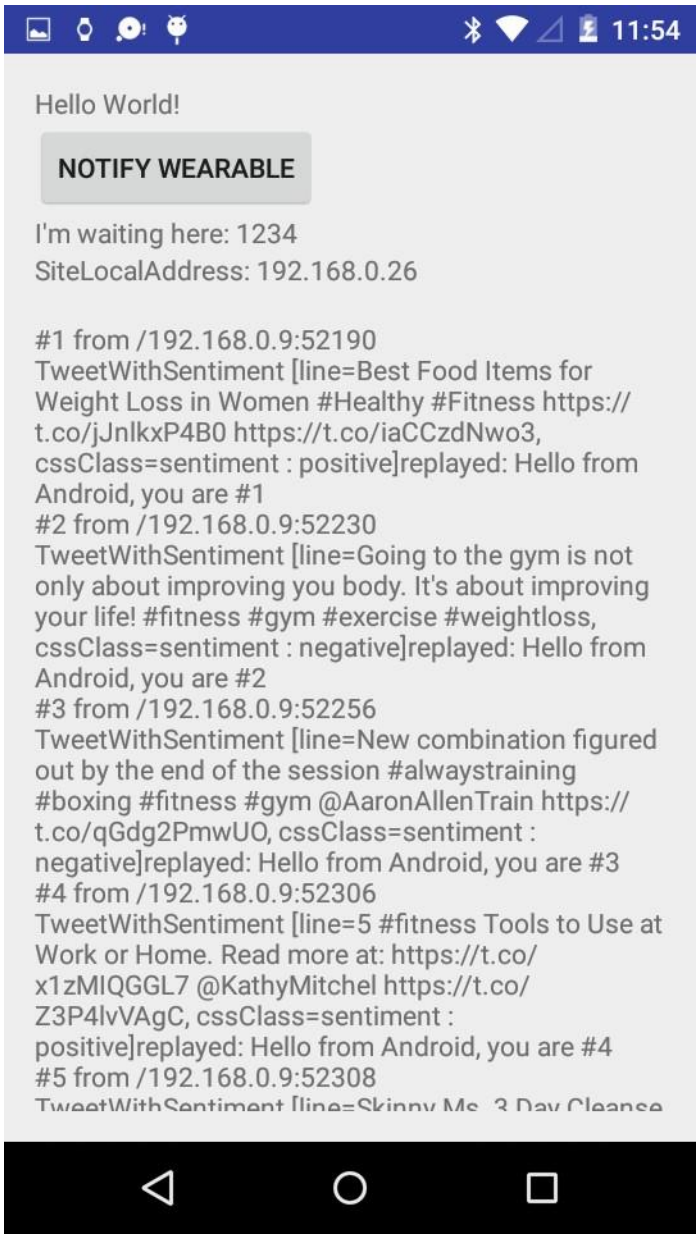
This is smartphone-smartwatch application in which the smart watch senses the data and the data is collected, stored in cloud database. The analysis is performed on the collected data and the results are sent as an email or notification the smartphone. The Spark Mlib and machine learning algorithms are used to perform analysis and determine the patterns from the user health conditions.

Machine Learning Application:

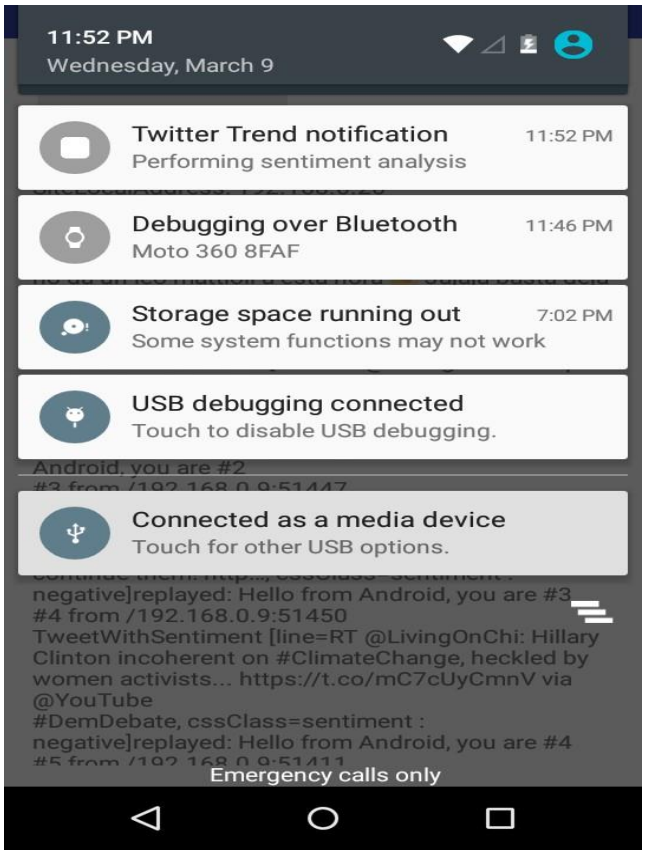
This is the main part of the application, where the machine analyzes and sends the suggests the patterns to user via mail. The machine learning algorithms provide high accuracy of data mining and results. It also results in different patterns that humans cannot determine at the same time. This is an application which is available in hand with the user and keeps a track of the medical history. This medical history can be used in medical field for research purposes as well.

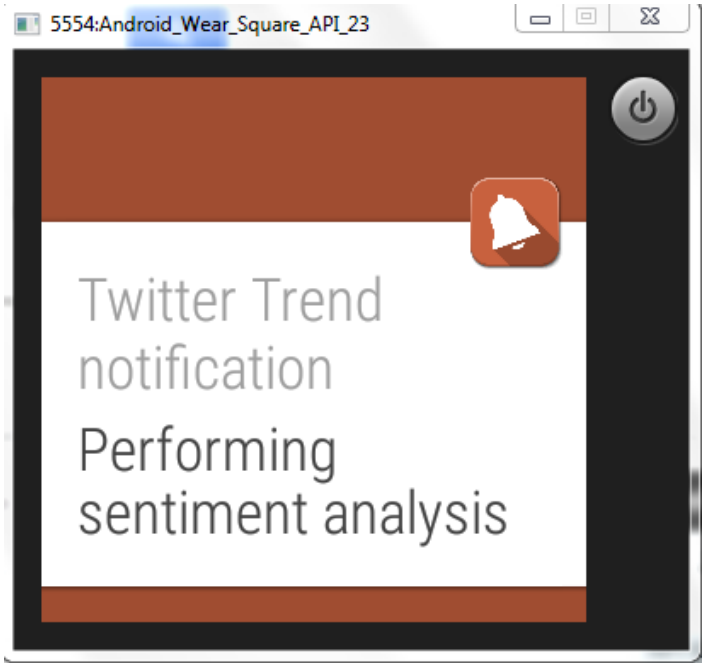
Sentimental Analysis:

The sentimental analysis is performed on the twitter live streaming data and the output is sent to the mobile through the socket connection. This would give us an idea of how the twitter tweets are being posted related to our project.



It is also sent as notification to the smart watch and also to the smartphone. The screenshots are placed below.



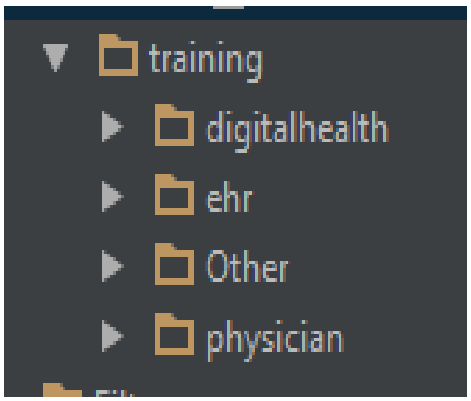


CLASSIFICATION:

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

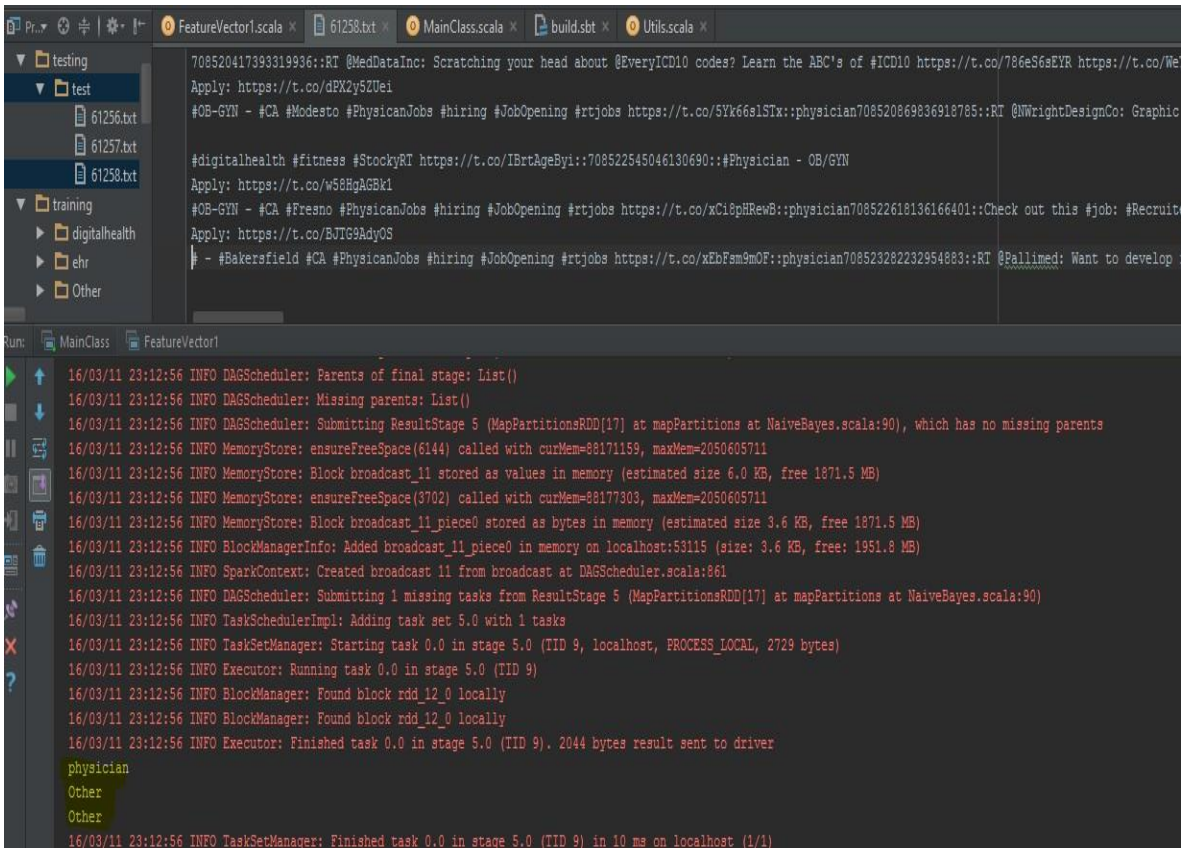
For now we have divided it into below mentioned 4 categories. These categories are on basis of hashtags:



We are training the data for this 4 categories and will test them against the future tweets. For now its just 4 categories but we will include more and plan is to ask user for his interest of tweet and on basis of that we will find appropriate tweet. The selection criteria for choosing tweet from streaming will be on basis of ratings. Program will decide ratings on basis of user's input and how influenced that tweet is to others people so far. We will include sentiment analysis as well for making sure that positive tweets will reach to user.
Below is the sample predicted results for few tweets.

61257.txt 61258.txt 61256.txt

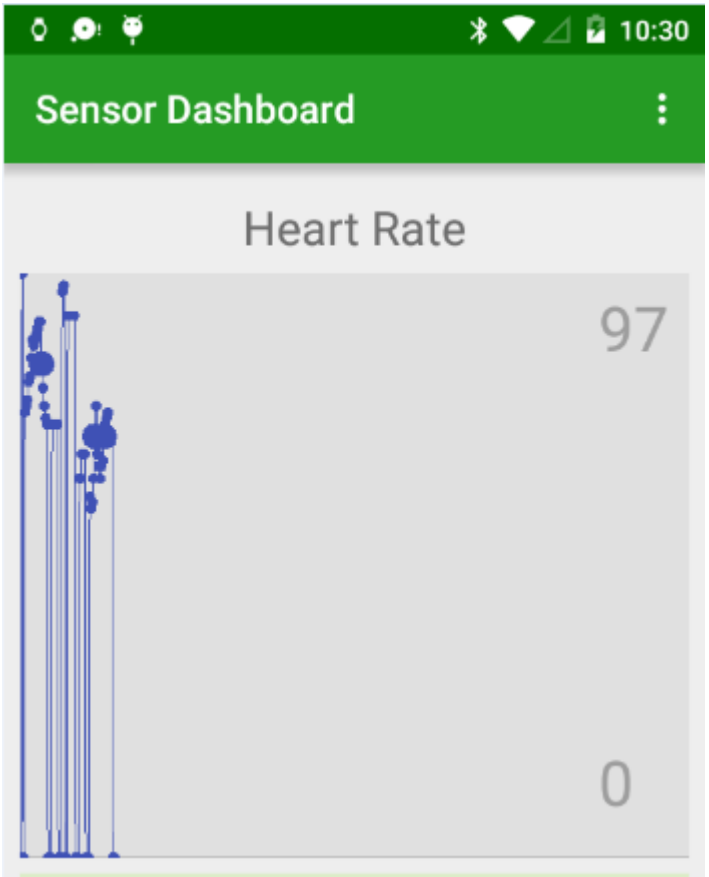
Above three files are the collected tweets during streaming and below is the predicted output:


FeatureVector1.scala
61258.txt
MainClass.scala
build.sbt
Utils.scala
testing
test
61256.txt
61257.txt
61258.txt
training
digitalhealth
ehr
Other
708520417393319936::RT @MedDataInc: Scratching your head about @EveryICD10 codes? Learn the ABC's of #ICD10 https://t.co/786eS6sEYR https://t.co/We
Apply: https://t.co/dPX2y5ZUei
#OB-GYN - #CA #Modesto #PhysicanJobs #hiring #JobOpening #rtjobs https://t.co/5Yk66slSTx::physician708520869836918785::RT @NWrightDesignCo: Graphic
#digitalhealth #fitness #StockyRT https://t.co/IBrtAgeByi::708522545046130690::#Physician - OB/GYN
Apply: https://t.co/w58HgAGBk1
#OB-GYN - #CA #Fresno #PhysicanJobs #hiring #JobOpening #rtjobs https://t.co/xCi8pHRewB::physician708522618136166401::Check out this #job: #Recruit
Apply: https://t.co/BJTG9AdyOS
- #Bakersfield #CA #PhysicanJobs #hiring #JobOpening #rtjobs https://t.co/xEbFsm9mOF::physician708523282232954883::RT @Pallimed: Want to develop
Run:
MainClass
FeatureVector1
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)

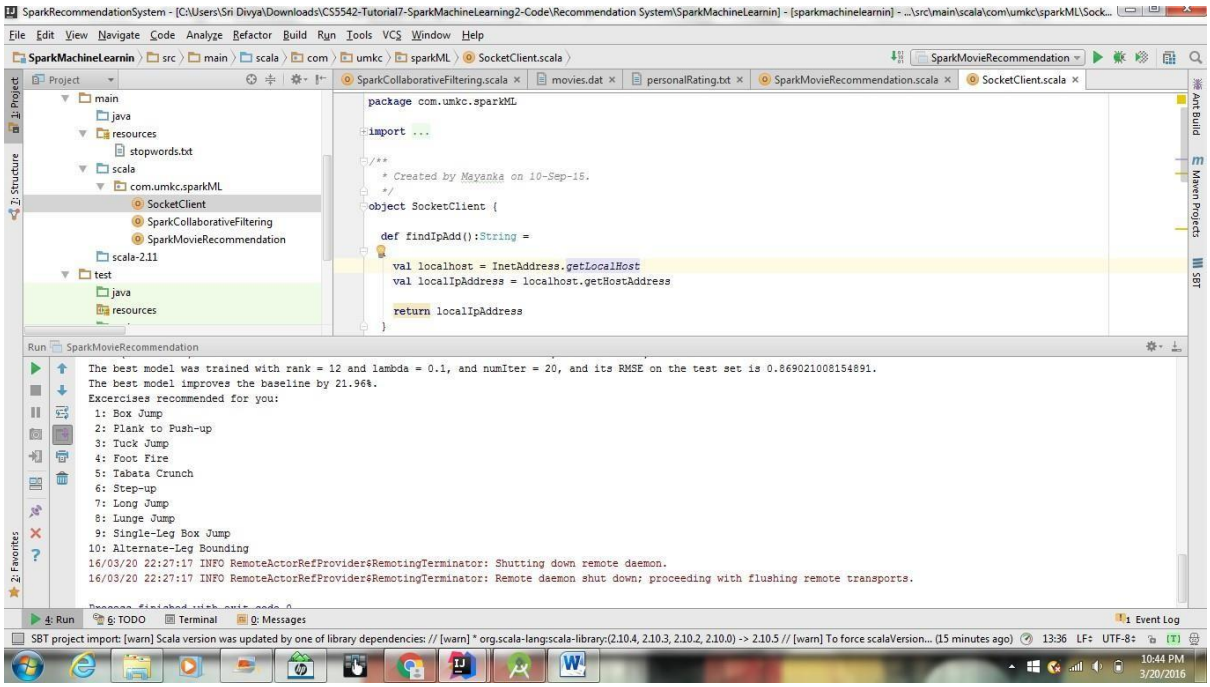
DEPLOYMENT:


Ultra.me
11:03
APPS
WIDGETS
Maps
Messaging
Messenger
Migrate
Notification
NotificationToSmartWatch
Phone
PhoneSensors
PhotoCollage
Play Books
Play Games
Play Movies & TV
Play Music
Play Newsstand
Play Store
Quickoffice
HeartFit application

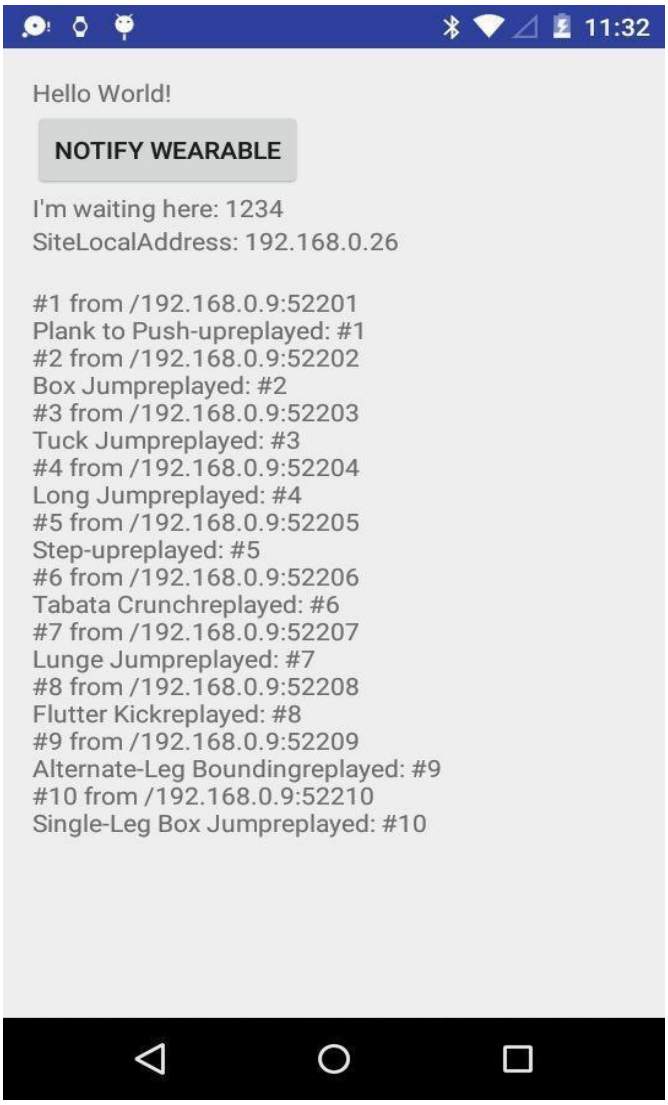

Steps : 113.0
HeartRate : 82.0


10:30
Sensor Dashboard
Heart Rate
97
0

In our project, we are recommending the users some physical and heart exercises based on their health condition and their daily routine. Based on the personal ratings and ratings of other users followed the recommendations are given.



The same is sent to smartwatch/ smart phone.



Frequent Pattern Mining – Spark - MLib

For matching the frequent patterns , we have used FP growth alogorithm.

FP-growth

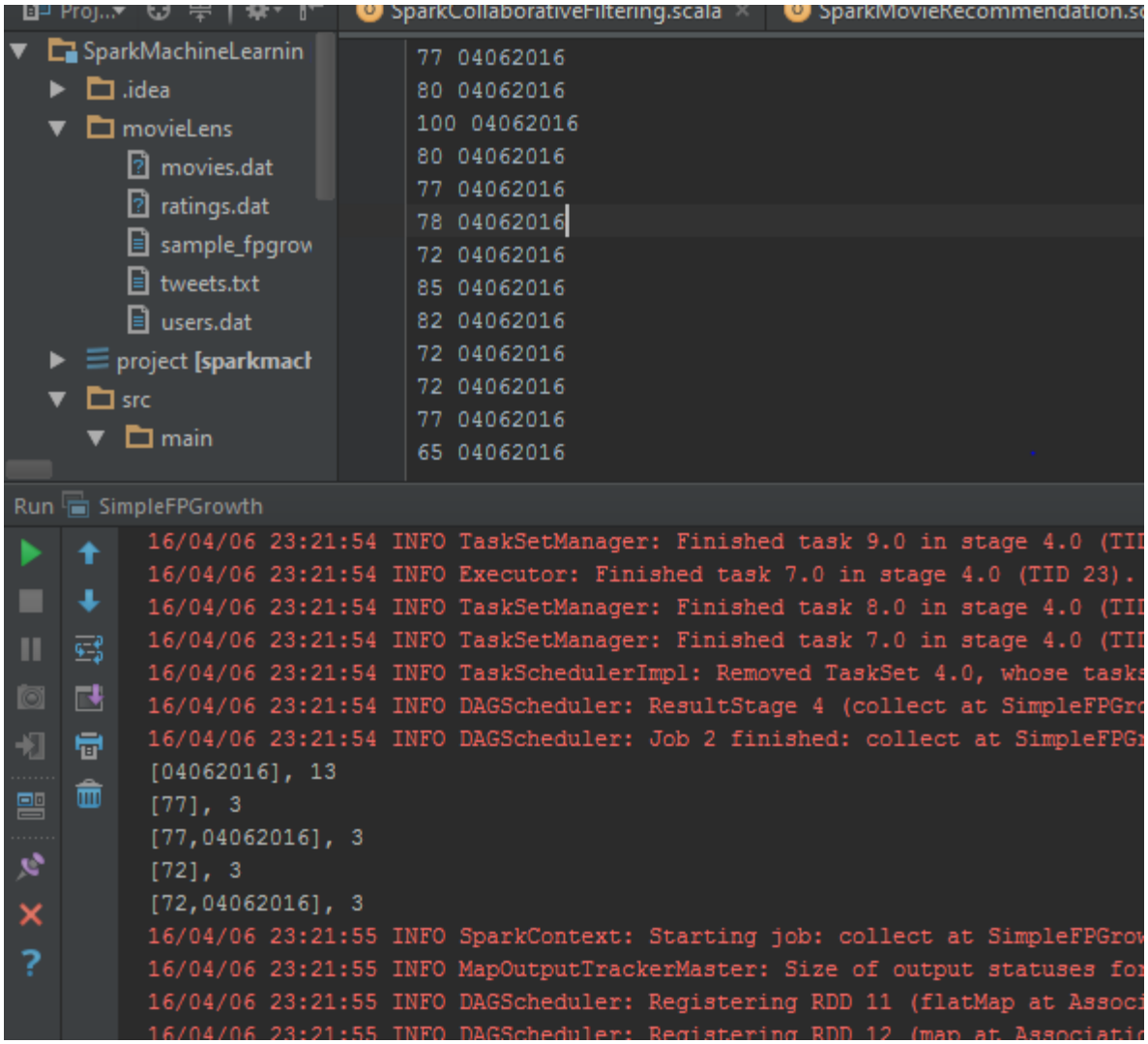
`spark.mllib`'s FP-growth implementation takes the following (hyper-)parameters:

- `minSupport`: the minimum support for an itemset to be identified as frequent. For example, if an item appears 3 out of 4 transactions, it has a support of 3/4=0.75.
- `numPartitions`: the number of partitions that are used to distribute the work.

Below is the implementation of algorithm for our heart bit data to find out the pattern on partocular day to check working of heart bit, whether its behaving normal or not. The output for our sample file is

included below it.

```scala
object SimpleFPGrowth {

  def main(args: Array[String]) {
    System.setProperty("hadoop.home.dir","c:\\winutils")

    val conf =new SparkConf().setMaster("local[*]").setAppName("SparkNaiveBayes").set("spark.driver.memory", "3g").set("spark.executor.mem
    val sc = new SparkContext(conf)
   // val conf = new SparkConf().setAppName("SimpleFPGrowth")
    //val sc = new SparkContext(conf)

    // $example on$
    val data = sc.textFile("movieLens/sample_fpgrowth.txt")

    val transactions: RDD[Array[String]] = data.map(s => s.trim.split(' '))

    val fpg = new FPGrowth()
      .setMinSupport(0.2)
      .setNumPartitions(10)
    val model = fpg.run(transactions)

    model.freqItemsets.collect().foreach { itemset =>
      println(itemset.items.mkString("[", ",", "]") + ", " + itemset.freq)
    }

    val minConfidence = 0.8
    model.generateAssociationRules(minConfidence).collect().foreach { rule =>
      println(
        rule.antecedent.mkString("[", ",", "]")
          + " => " + rule.consequent .mkString("[", ",", "]")
          + ", " + rule.confidence)
    }
```



Github link: https://github.com/hirenshah7390/Bigdata_Project/tree/master/Hiren

PROJECT MANAGEMENT:

Planning

We as a team discussed about the project idea, project flow , features that are to be implemented. Roles and responsibilities are being discussed and given below.

Time: 8 hours

Members Participated: HirenShah, Abhiram, Harshini, Dinesh Reddy

Implementation

The step counter and heart sensor data is collected continuously using the sensors in the smart watch.

The data is sent to the smartphone and graphs are plotted on phone. The collected data is classified and

working on recommendation systems.

The machine learning algorithms were implemented and the recommendation systems was implemented

Responsibility: Data collection, Zenhub, FP growth algorithm, Recommendation system.

Time: 30 hours

Participants: Harshini, Abhiram, HirenShah, Dinesh Reddy

Testing

Test cases for all the above designed pages were implemented.

Responsibility: Tried to collect data at different times when sleeping, walking,etc

Time: 4 hours

Participants: HirenShah, Dinesh Reddy, Abhiram, Harshini

BIBILIOGRAPHY:

http://www.r-bloggers.com/association-rule-learning-and-the-apriori-algorithm/

http://www.dreamincode.net/forums/topic/324137-periodically-collect-accelerometer-data-in-android/

http://www.wareable.com/fitbit/fitbit-70-of-people-ignore-heart-rate-data-1523

https://dev.fitbit.com/docs/heart-rate/

http://www.livescience.com/42132-heart-rate-activity-tracker-useful.html

http://spark.apache.org/docs/latest/mllib-guide.html