USER ROUTINE PREDICTIVE ANALYSIS

BIG DATA ANALYTICS AND APPLICATIONS PROJECT

SPRING 2016

SUBMITTED BY

RAKESH REDDY BANDI (02)

NIHAR DUDAM (06)

SRICHARAN PAMURU (20)

MARK SCHULTZ (26)

I. Introduction:

Big data has become a buzzword in the field of Computer Science. As technology advances, we are finding new ways to collect and analyze data. With the invention of the smart watch, we have been able to expand this collection of data to areas that had not been possible years ago. In addition, it has never been easier for the user to retrieve data from their device when needed. Our project combines these two conveniences into one application. We will use big data to measure and map a User's routine, and use this routine to make predictions and suggestions about the future.

We plan on applying several sources for data collection. First of all, the applications collect data from movements of the watch to classify patterns in certain movements. It will use these movements to guess the activity that the User may be doing at the moment. It will also collect data from the heart rate monitor of the smartwatch to measure the rigor of the User's activity. GPS data on the watch will determine the location of the user. We will also use datetime metrics to determine the day and the time that the user typically does certain activities.

The sources of data above will be collected by the user, but we also want to combine this with several API's to make better predictions, and add suggestions to give back to the user. We will use speech to text API to query the application for information. We will use the Google Location API to develop a richer context of the user's most frequented locations, and suggest possible locations that might interest the user. We will implement the Yelp API, so that when the application predicts when the user eats, it will make suggestions on places to go.

All of the data we will collect will be organized using the Hadoop file system and implemented using Amazon Web Services. We will discuss the technologies more in this documentation. The project is broken down into four phases. For the first Increment, we began the data collection to measure several common movements. We are initially using Naïve Bayes to classify these movements. We have also began research on several of the API's that we plan on implementing.

II. Project Goal and Objectives:

The main implementation of the project will consist of the following specification based objective and subjective analysis along with a sketched perspective of goal:

1. Overall Goal:

The goal of the project is to predict a user's activity at any time during a user's schedule. A large dataset is provided by the smartwatch and the machine is trained on the basis of machine learning algorithms as to how the coordinates will affect the motion of the sensor. Machine is trained and the provided data is analyzed based on the reading from the various sensors of the smartwatch. The location information will provide a set of variable features such as restaurants, parks, etc. near to the naive user. APIs will provide the analysis of information and data regarding the review and further circumspect options are brought forward to help the smartwatch analyze the data and provide a predictive routine. A variable set of sensors will collect the proximity, heart rate, accelerometer based motion readings, etc. to analyze data set. The analysis of this large datasets will return recommended notifications to the user for a viable routine.

2. Specific objectives:

- Predict user routine by analysis of large datasets through machine learning techniques.
- Return geolocation based places in proximity such as restaurants, parks, etc.
- Sensor dataset provides specific activity based readings.
- Provide recommended services by sending notification to user depending on time and routine.

3. Specific features:

Sensors in the watch serve as primary source for providing data which is required for analysis. The accelerometer will provide a three dimensional reading with regard to the position. The heart rate sensor will provide the heart beat rate which can be used to obtain the magnitude of force exerted while performing a particular activity. GPS will provide information about the location and the APIS will return data to the user depending upon the routine's requirements and suggestive notifications will be sent to the smartwatch while the connection to another remote device or machine such as a smartphone holds true. It will create a complicate alliance and thus the machine is to be trained using various machine learning algorithms. Data clusters are classified using the classification techniques intended to provide more information about the routine prediction. There are other sensors in the watch such as ambient light sensor, gyroscope, magnetic field sensor, step counter, gravity, linear acceleration, significant motion, etc. which are used collaboratively to obtain a dataset.

4. Significance:

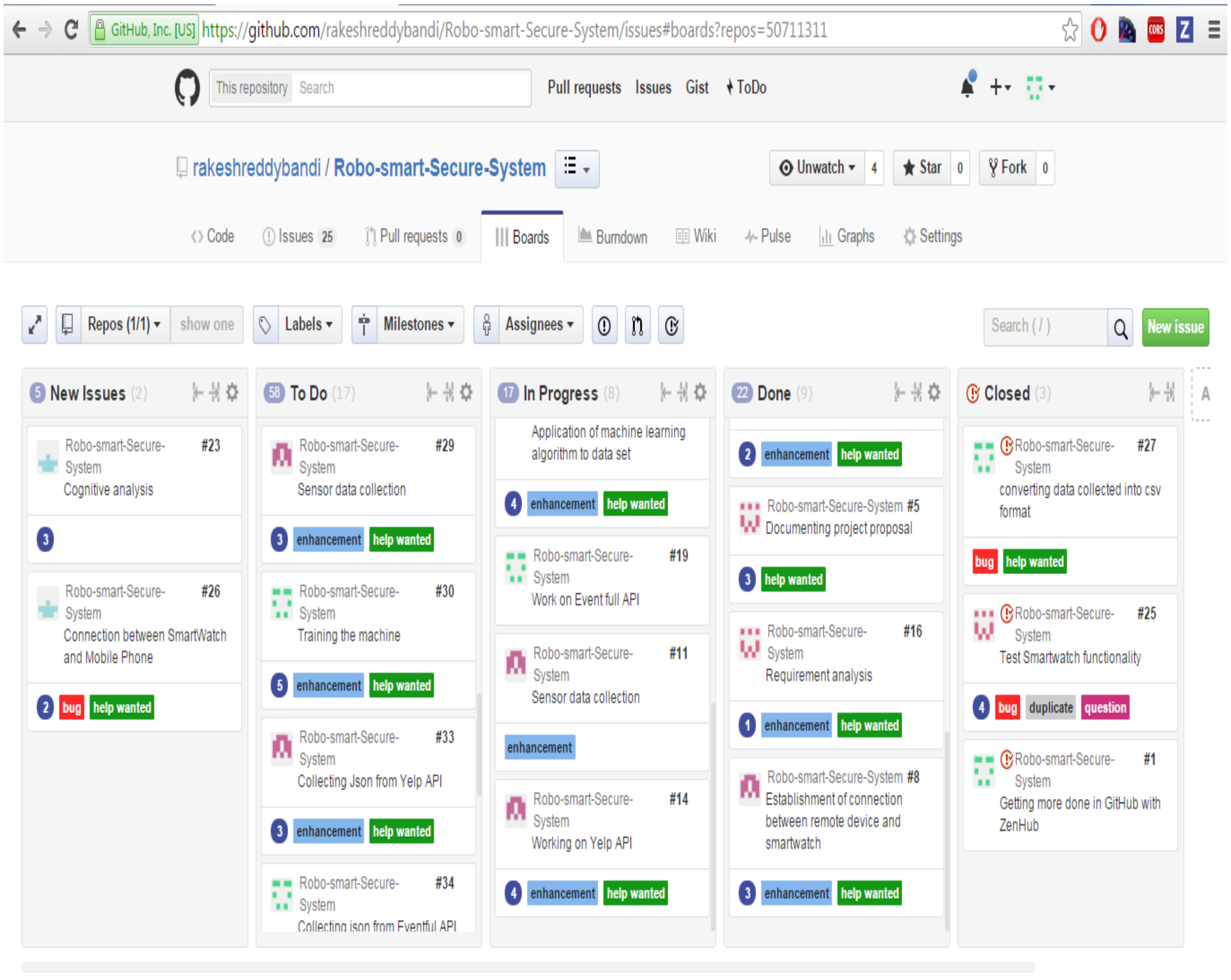
The most important aspect of the project is that an analysis of big data is done in order to predict the routine of a specific user. It will generate a pattern based on the different readings to generate a more sophisticated clustered activity set. This will help in obtaining a clear cut view of a user's routine plus a more generative dimension to a time schedule. Sensors present in the smartwatch will be crucial in collection of data, which in turn is fed into the machine through various machine learning techniques. A stable schedule would prompt the user to receive a collaborative approach to obtain the output as a predictive routine analysis.

III. Project Plan:

1. Schedule for entire project:

The entire schedule for the project is shown as follows where the respective use case and process step up with all the modules and the time frame is categorized specifically to illustrate the step by step fulfilment of the entire processes done such as feature design and feature implementation.

Screenshot 1: The following screenshot demonstrates the overall perspective of implementation in which the project is being done. The To do and In Progress have a significantly high number of modules and are continued in the following images.



GitHub, Inc. [US] https://github.com/rakeshreddybandi/Robo-smart-Secure-System/issues#boards?repos=50711311
This repository Search
Pull requests Issues Gist ToDo
rakeshreddybandi / Robo-smart-Secure-System
Unwatch 4 Star 0 Fork 0
Code Issues 25 Pull requests 0 Boards Burndown Wiki Pulse Graphs Settings
Repos (1/1) show one Labels Milestones Assignees
Search (/) New issue

5 New Issues (2)
- Robo-smart-Secure-System #23 Cognitive analysis — 3
- Robo-smart-Secure-System #26 Connection between SmartWatch and Mobile Phone — 2 bug help wanted

58 To Do (17)
- Robo-smart-Secure-System #29 Sensor data collection — 3 enhancement help wanted
- Robo-smart-Secure-System #30 Training the machine — 5 enhancement help wanted
- Robo-smart-Secure-System #33 Collecting Json from Yelp API — 3 enhancement help wanted
- Robo-smart-Secure-System #34 Collecting json from Eventful API

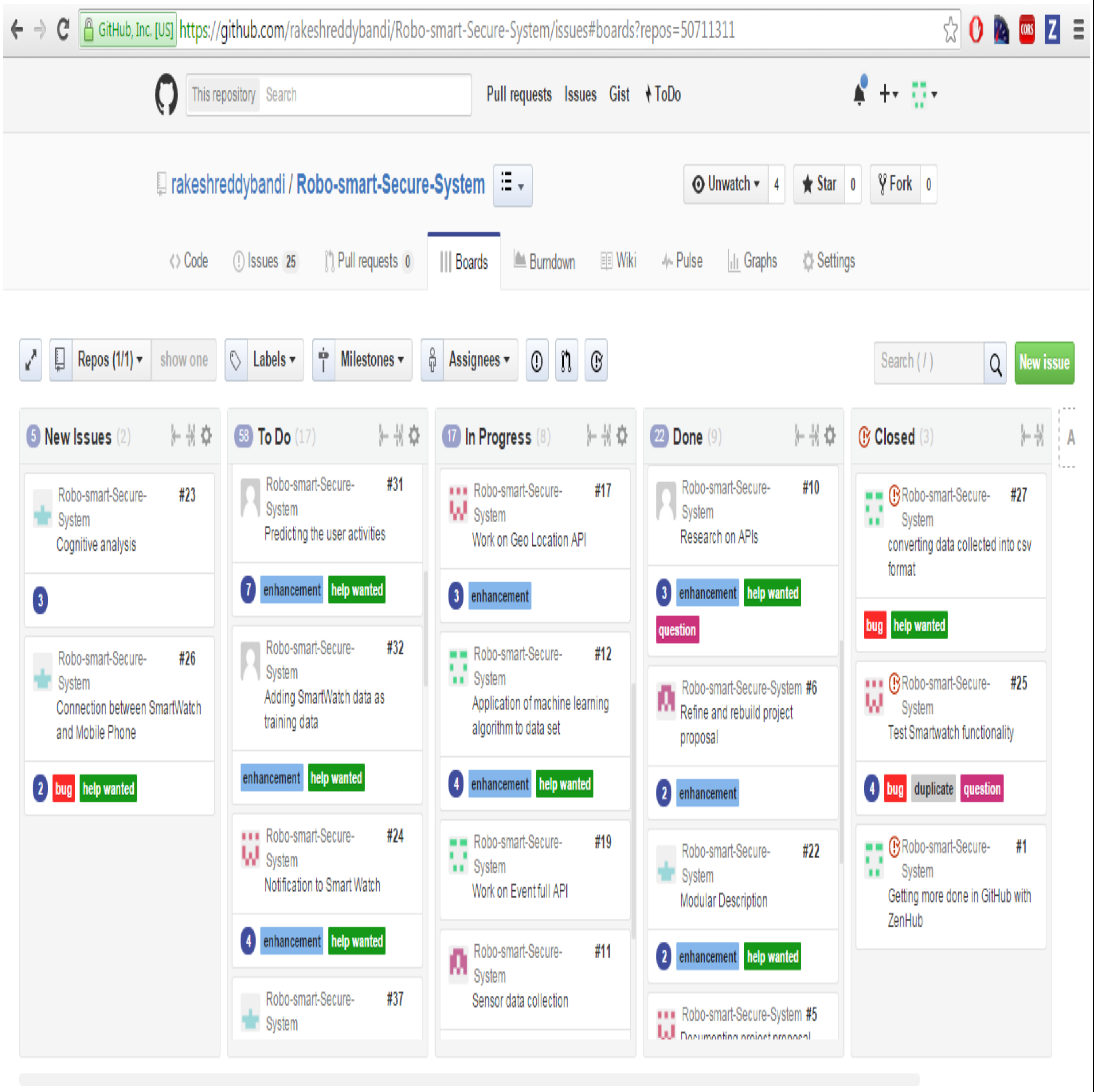
17 In Progress (8)
- Application of machine learning algorithm to data set — 4 enhancement help wanted
- Robo-smart-Secure-System #19 Work on Event full API
- Robo-smart-Secure-System #11 Sensor data collection — enhancement
- Robo-smart-Secure-System #14 Working on Yelp API — 4 enhancement help wanted

22 Done (9)
- 2 enhancement help wanted
- Robo-smart-Secure-System #5 Documenting project proposal — 3 help wanted
- Robo-smart-Secure-System #16 Requirement analysis — 1 enhancement help wanted
- Robo-smart-Secure-System #8 Establishment of connection between remote device and smartwatch — 3 enhancement help wanted

Closed (3)
- Robo-smart-Secure-System #27 converting data collected into csv format — bug help wanted
- Robo-smart-Secure-System #25 Test Smartwatch functionality — 4 bug duplicate question
- Robo-smart-Secure-System #1 Getting more done in GitHub with ZenHub

Screenshot 2:

The following screenshots shows the second listing of the To Do phases and In Progress where there is an increase in the number of elements and tasks to be done. There is an incentive in populating the respective tasks and assign them to the members as per core requirement and generative design.


https://github.com/rakeshreddybandi/Robo-smart-Secure-System/issues#boards?repos=50711311
This repository Search
Pull requests Issues Gist ToDo
rakeshreddybandi / Robo-smart-Secure-System
Unwatch 4 Star 0 Fork 0
Code Issues 25 Pull requests 0 Boards Burndown Wiki Pulse Graphs Settings
Repos (1/1) show one Labels Milestones Assignees
Search (/) New issue

New Issues (2)
- Robo-smart-Secure-System #23 Cognitive analysis — 3
- Robo-smart-Secure-System #26 Connection between SmartWatch and Mobile Phone — 2 bug, help wanted

To Do (17)
- Robo-smart-Secure-System #31 Predicting the user activities — 7 enhancement, help wanted
- Robo-smart-Secure-System #32 Adding SmartWatch data as training data — enhancement, help wanted
- Robo-smart-Secure-System #24 Notification to Smart Watch — 4 enhancement, help wanted
- Robo-smart-Secure-System #37

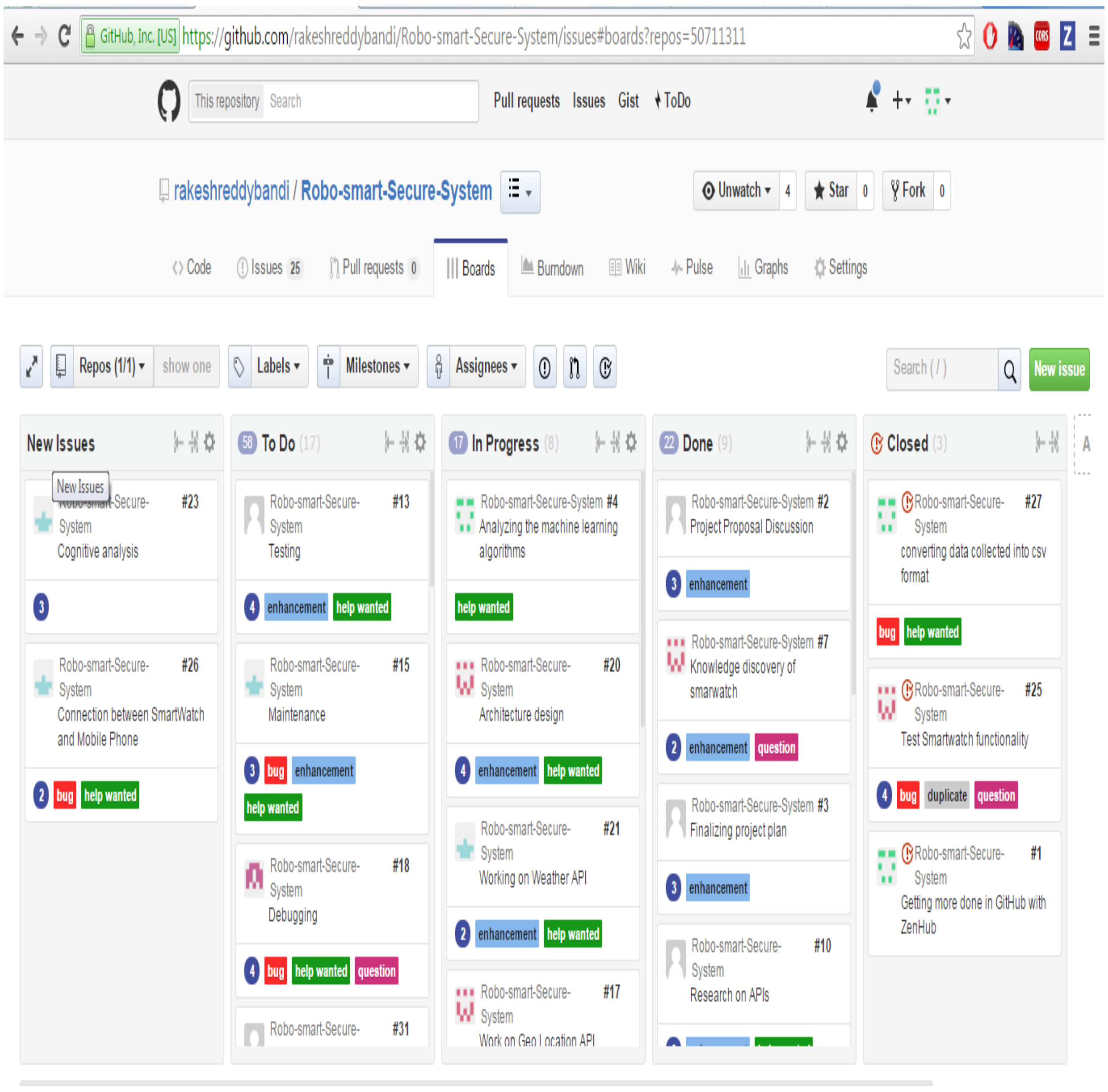
In Progress (8)
- Robo-smart-Secure-System #17 Work on Geo Location API — 3 enhancement
- Robo-smart-Secure-System #12 Application of machine learning algorithm to data set — 4 enhancement, help wanted
- Robo-smart-Secure-System #19 Work on Event full API
- Robo-smart-Secure-System #11 Sensor data collection

Done (9)
- Robo-smart-Secure-System #10 Research on APIs — 3 enhancement, help wanted, question
- Robo-smart-Secure-System #6 Refine and rebuild project proposal — 2 enhancement
- Robo-smart-Secure-System #22 Modular Description — 2 enhancement, help wanted
- Robo-smart-Secure-System #5 Documenting project proposal

Closed (3)
- Robo-smart-Secure-System #27 converting data collected into csv format — bug, help wanted
- Robo-smart-Secure-System #25 Test Smartwatch functionality — 4 bug, duplicate, question
- Robo-smart-Secure-System #1 Getting more done in GitHub with ZenHub

Screenshot 3:

The following screenshot shows the phase where the tasks in progress have been concentrated mainly on the design specifications. It combines along with the done tasks to stabilize the development of project in due course to simultaneously bring out the working modules.



GitHub, Inc. [US] https://github.com/rakeshreddybandi/Robo-smart-Secure-System/issues#boards?repos=50711311

This repository Search Pull requests Issues Gist ToDo

rakeshreddybandi / Robo-smart-Secure-System — Unwatch 4 · Star 0 · Fork 0

Code · Issues 25 · Pull requests 0 · Boards · Burndown · Wiki · Pulse · Graphs · Settings

Repos (1/1) · show one · Labels · Milestones · Assignees · Search (/) · New issue

New Issues
- Robo-smart-Secure-System #23 — Cognitive analysis — 3
- Robo-smart-Secure-System #26 — Connection between SmartWatch and Mobile Phone — 2 bug, help wanted

58 To Do (17)
- Robo-smart-Secure-System #13 — Testing — 4 enhancement, help wanted
- Robo-smart-Secure-System #15 — Maintenance — 3 bug, enhancement, help wanted
- Robo-smart-Secure-System #18 — Debugging — 4 bug, help wanted, question
- Robo-smart-Secure- #31

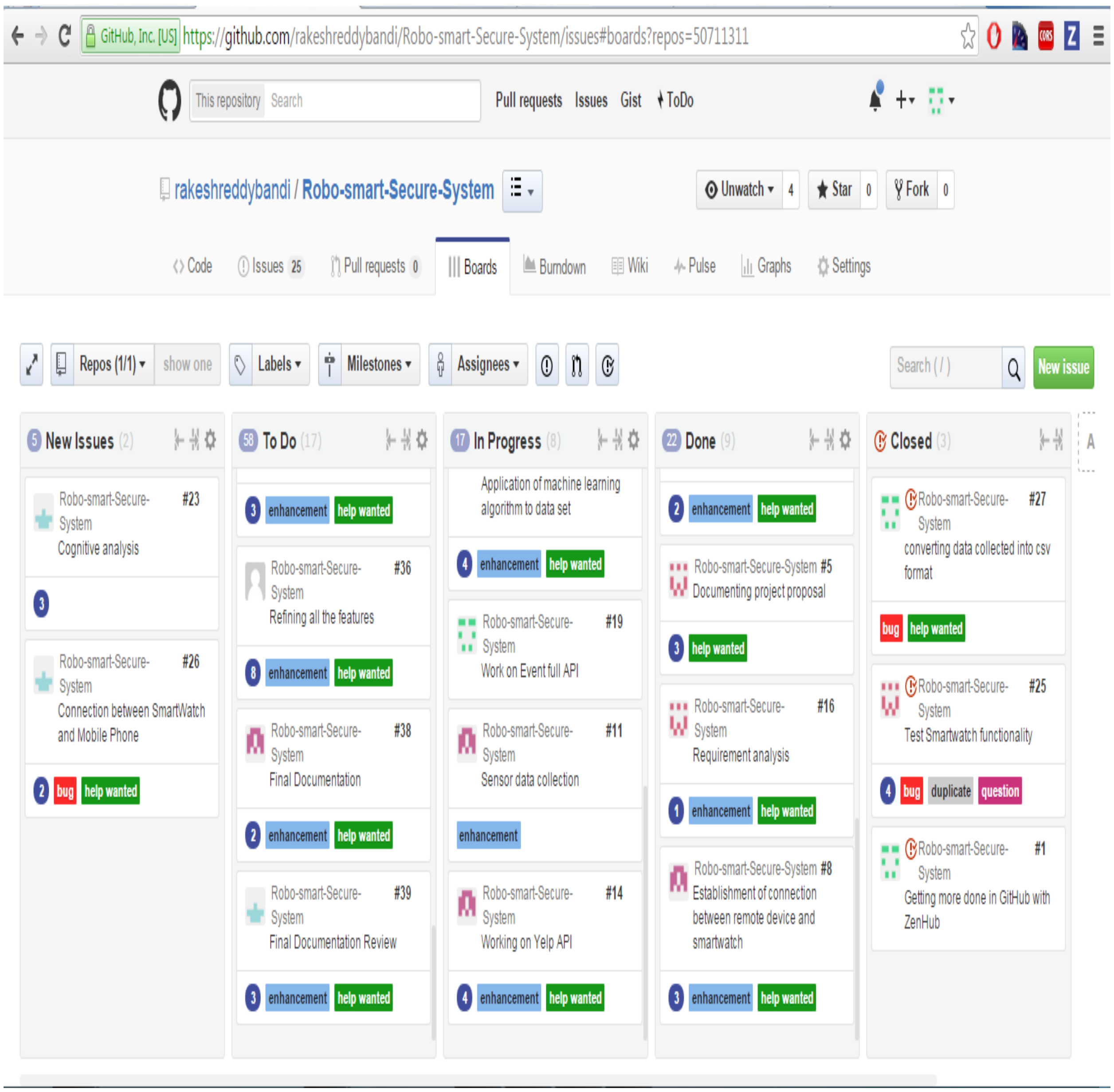
17 In Progress (8)
- Robo-smart-Secure-System #4 — Analyzing the machine learning algorithms — help wanted
- Robo-smart-Secure-System #20 — Architecture design — 4 enhancement, help wanted
- Robo-smart-Secure-System #21 — Working on Weather API — 2 enhancement, help wanted
- Robo-smart-Secure-System #17 — Work on Geo Location API

22 Done (9)
- Robo-smart-Secure-System #2 — Project Proposal Discussion — 3 enhancement
- Robo-smart-Secure-System #7 — Knowledge discovery of smarwatch — 2 enhancement, question
- Robo-smart-Secure-System #3 — Finalizing project plan — 3 enhancement
- Robo-smart-Secure-System #10 — Research on APIs

Closed (3)
- Robo-smart-Secure-System #27 — converting data collected into csv format — bug, help wanted
- Robo-smart-Secure-System #25 — Test Smartwatch functionality — 4 bug, duplicate, question
- Robo-smart-Secure-System #1 — Getting more done in GitHub with ZenHub

Screenshot 4:

The screenshot here shows the listing done where the penultimate requirement analysis and engineering is done to bring out the resulted output that is tested. The maintenance is to be done and especially debugging to negate results which are not in favour of project deployment.



GitHub, Inc. [US] https://github.com/rakeshreddybandi/Robo-smart-Secure-System/issues#boards?repos=50711311

This repository Search Pull requests Issues Gist ToDo

rakeshreddybandi / Robo-smart-Secure-System Unwatch 4 Star 0 Fork 0

Code Issues 25 Pull requests 0 Boards Burndown Wiki Pulse Graphs Settings

Repos (1/1) show one Labels Milestones Assignees Search (/) New issue

5 New Issues (2)

Robo-smart-Secure-System #23
Cognitive analysis
3

Robo-smart-Secure-System #26
Connection between SmartWatch and Mobile Phone
2 bug help wanted

58 To Do (17)

3 enhancement help wanted

Robo-smart-Secure-System #36
Refining all the features
8 enhancement help wanted

Robo-smart-Secure-System #38
Final Documentation
2 enhancement help wanted

Robo-smart-Secure-System #39
Final Documentation Review
3 enhancement help wanted

17 In Progress (8)

Application of machine learning algorithm to data set
4 enhancement help wanted

Robo-smart-Secure-System #19
Work on Event full API

Robo-smart-Secure-System #11
Sensor data collection
enhancement

Robo-smart-Secure-System #14
Working on Yelp API
4 enhancement help wanted

22 Done (9)

2 enhancement help wanted

Robo-smart-Secure-System #5
Documenting project proposal
3 help wanted

Robo-smart-Secure-System #16
Requirement analysis
1 enhancement help wanted

Robo-smart-Secure-System #8
Establishment of connection between remote device and smartwatch
3 enhancement help wanted

Closed (3)

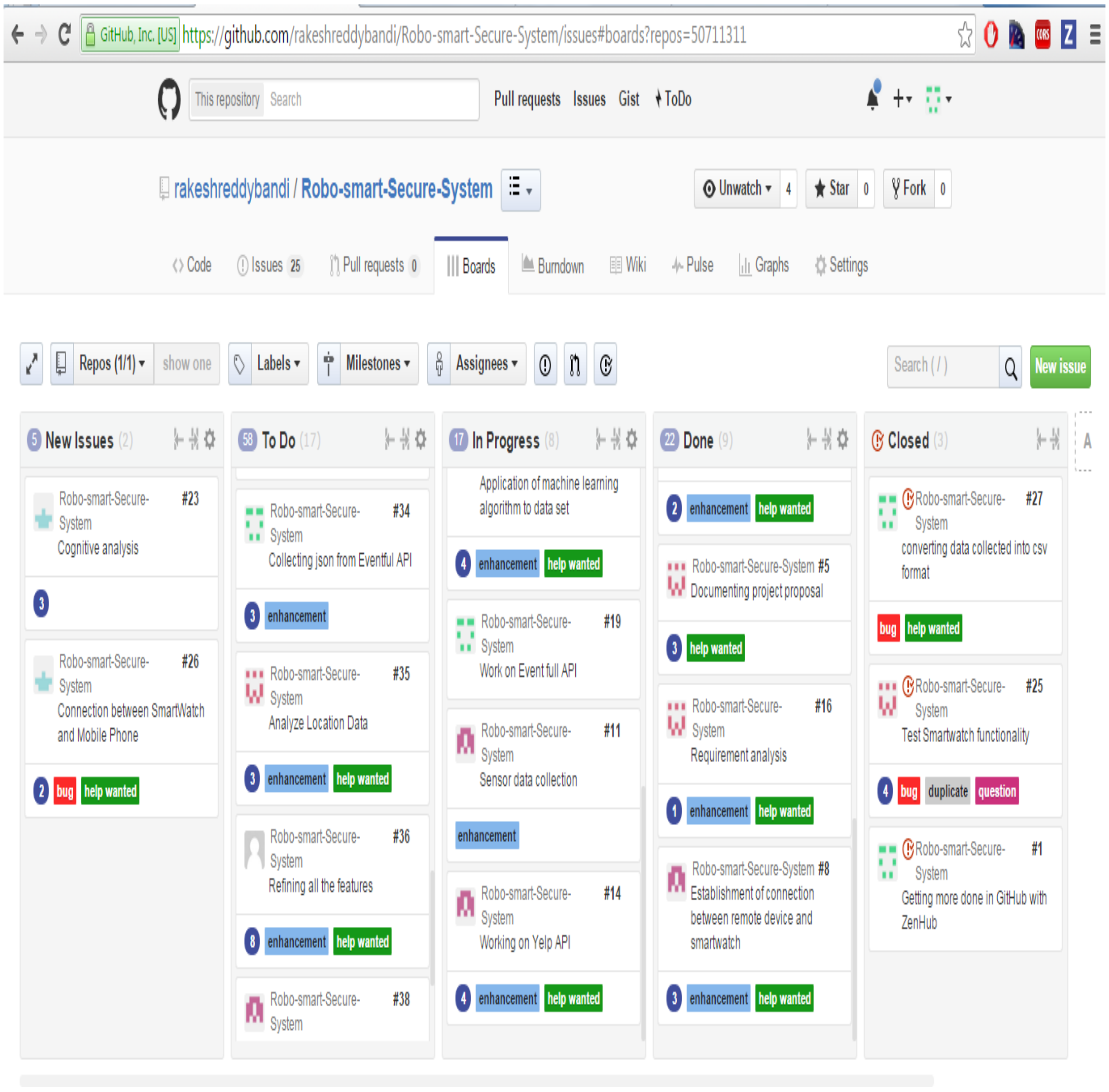
Robo-smart-Secure-System #27
converting data collected into csv format
bug help wanted

Robo-smart-Secure-System #25
Test Smartwatch functionality
4 bug duplicate question

Robo-smart-Secure-System #1
Getting more done in GitHub with ZenHub

Screenshot 5:

The entire project schedule has been roughly shown here. A detailed insight of project plan is done in such a manner that there might be few more additional development or phase build in due course of the project. It shows a correlation of the various modules included as per specification and features.



GitHub, Inc. [US] https://github.com/rakeshreddybandi/Robo-smart-Secure-System/issues#boards?repos=50711311

rakeshreddybandi / Robo-smart-Secure-System

Unwatch 4 · Star 0 · Fork 0

Code · Issues 25 · Pull requests 0 · Boards · Burndown · Wiki · Pulse · Graphs · Settings

New Issues (2)
- Robo-smart-Secure-System #23 — Cognitive analysis — 3
- Robo-smart-Secure-System #26 — Connection between SmartWatch and Mobile Phone — 2 bug, help wanted

To Do (17)
- Robo-smart-Secure-System #34 — Collecting json from Eventful API — 3 enhancement
- Robo-smart-Secure-System #35 — Analyze Location Data — 3 enhancement, help wanted
- Robo-smart-Secure-System #36 — Refining all the features — 8 enhancement, help wanted
- Robo-smart-Secure-System #38

In Progress (8)
- Application of machine learning algorithm to data set — 4 enhancement, help wanted
- Robo-smart-Secure-System #19 — Work on Event full API
- Robo-smart-Secure-System #11 — Sensor data collection — enhancement
- Robo-smart-Secure-System #14 — Working on Yelp API — 4 enhancement, help wanted

Done (9)
- 2 enhancement, help wanted
- Robo-smart-Secure-System #5 — Documenting project proposal — 3 help wanted
- Robo-smart-Secure-System #16 — Requirement analysis — 1 enhancement, help wanted
- Robo-smart-Secure-System #8 — Establishment of connection between remote device and smartwatch — 3 enhancement, help wanted

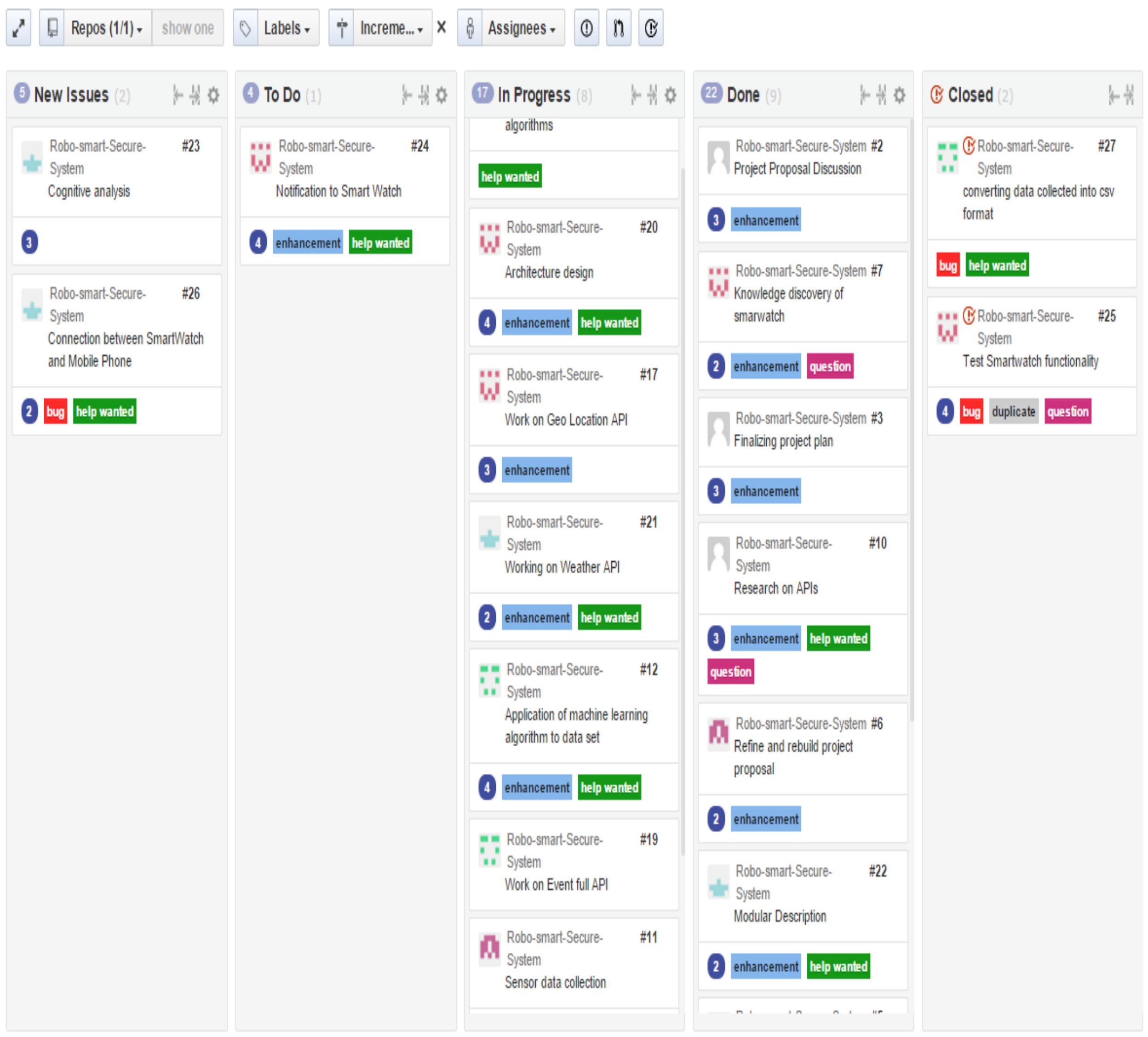
Closed (3)
- Robo-smart-Secure-System #27 — converting data collected into csv format — bug, help wanted
- Robo-smart-Secure-System #25 — Test Smartwatch functionality — 4 bug, duplicate, question
- Robo-smart-Secure-System #1 — Getting more done in GitHub with ZenHub

2. Project Timeline:

The implementation of the project is covered with a specific timeline and this is shown as follows to implement it using Zenhub:

Increment 1:

The first increment has the following tasks to be completed by 02/19/2016 with tasks assigned as such:



Repos (1/1) show one Labels Increme... × Assignees

5 New Issues (2)
- Robo-smart-Secure-System #23 — Cognitive analysis — 3
- Robo-smart-Secure-System #26 — Connection between SmartWatch and Mobile Phone — 2 bug, help wanted

4 To Do (1)
- Robo-smart-Secure-System #24 — Notification to Smart Watch — 4 enhancement, help wanted

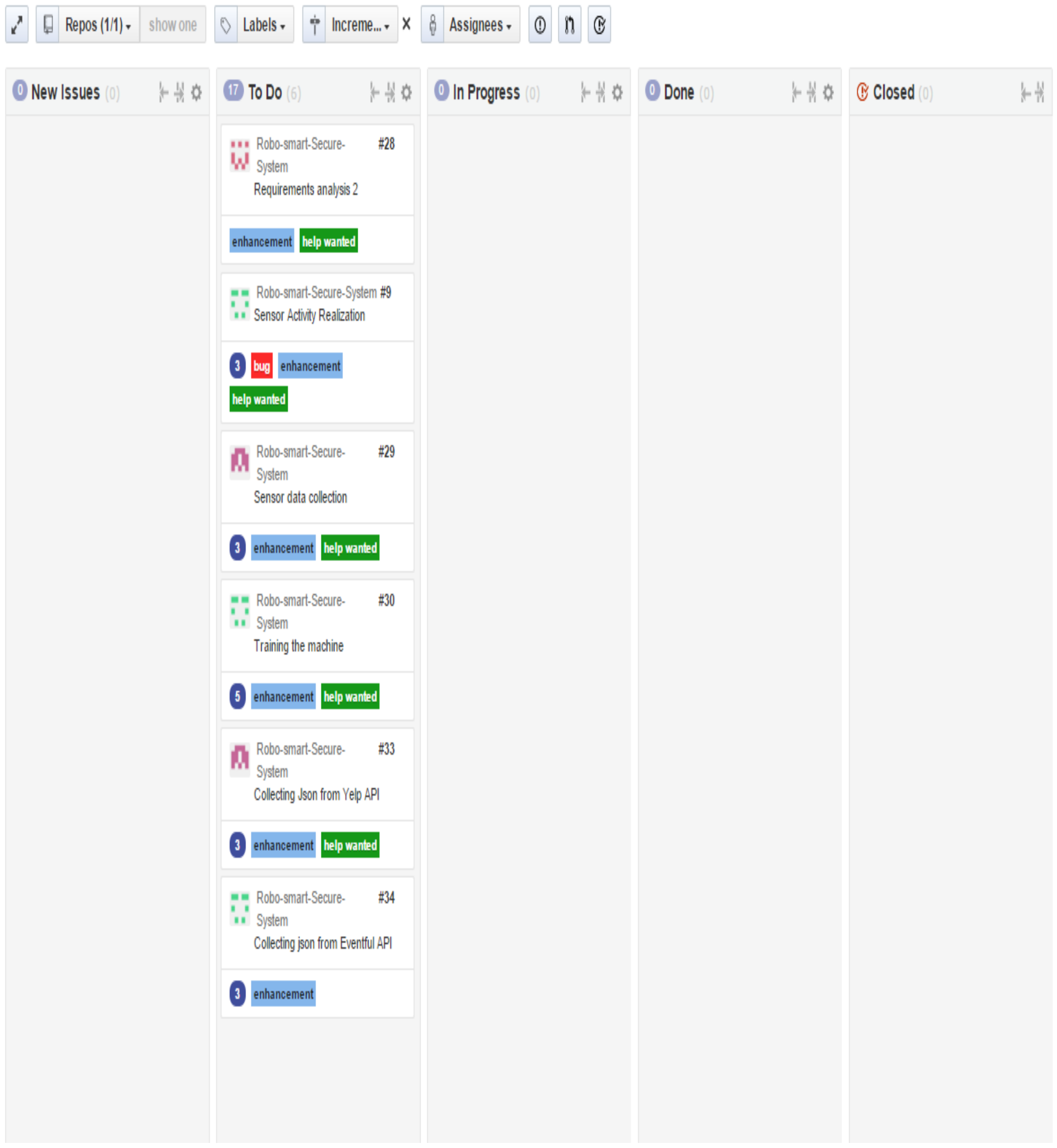
17 In Progress (8)
- algorithms — help wanted
- Robo-smart-Secure-System #20 — Architecture design — 4 enhancement, help wanted
- Robo-smart-Secure-System #17 — Work on Geo Location API — 3 enhancement
- Robo-smart-Secure-System #21 — Working on Weather API — 2 enhancement, help wanted
- Robo-smart-Secure-System #12 — Application of machine learning algorithm to data set — 4 enhancement, help wanted
- Robo-smart-Secure-System #19 — Work on Event full API
- Robo-smart-Secure-System #11 — Sensor data collection

22 Done (9)
- Robo-smart-Secure-System #2 — Project Proposal Discussion — 3 enhancement
- Robo-smart-Secure-System #7 — Knowledge discovery of smarwatch — 2 enhancement, question
- Robo-smart-Secure-System #3 — Finalizing project plan — 3 enhancement
- Robo-smart-Secure-System #10 — Research on APIs — 3 enhancement, help wanted, question
- Robo-smart-Secure-System #6 — Refine and rebuild project proposal — 2 enhancement
- Robo-smart-Secure-System #22 — Modular Description — 2 enhancement, help wanted

Closed (2)
- Robo-smart-Secure-System #27 — converting data collected into csv format — bug, help wanted
- Robo-smart-Secure-System #25 — Test Smartwatch functionality — 4 bug, duplicate, question

Increment 2:

The timeline for the second increment due on 03/11/2016 will have the graph as follows:



Robo-smart-Secure-System #28
Requirements analysis 2
enhancement help wanted

Robo-smart-Secure-System #9
Sensor Activity Realization
3 bug enhancement help wanted

Robo-smart-Secure-System #29
Sensor data collection
3 enhancement help wanted

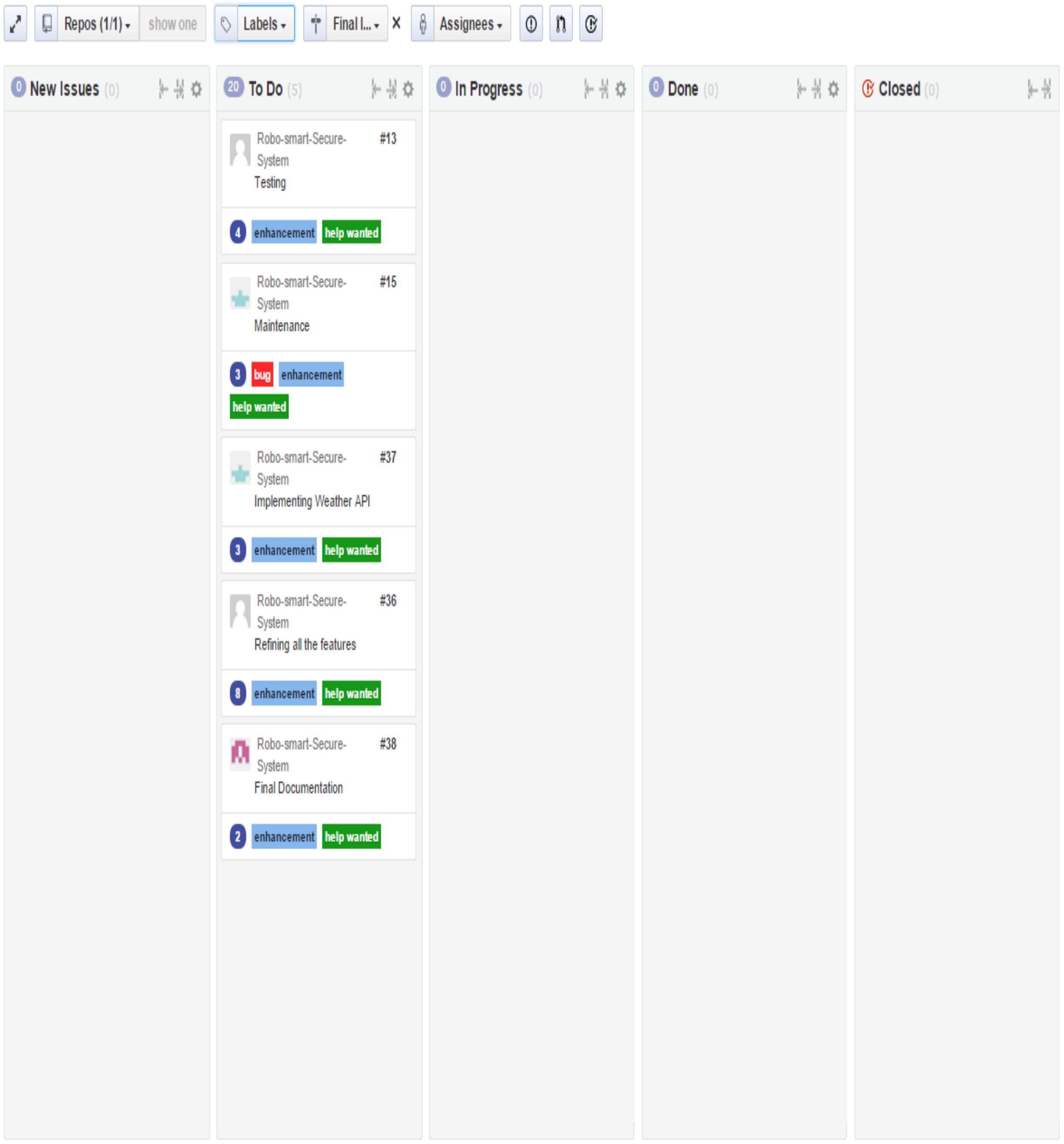
Robo-smart-Secure-System #30
Training the machine
5 enhancement help wanted

Robo-smart-Secure-System #33
Collecting Json from Yelp API
3 enhancement help wanted

Robo-smart-Secure-System #34
Collecting json from Eventful API
3 enhancement

Increment 3:

The timeline for the third increment due on 04/06/2016 will have the graph as follows:



0 New Issues (0)

20 To Do (5)

Robo-smart-Secure-System #13
Testing
4 enhancement help wanted

Robo-smart-Secure-System #15
Maintenance
3 bug enhancement help wanted

Robo-smart-Secure-System #37
Implementing Weather API
3 enhancement help wanted

Robo-smart-Secure-System #36
Refining all the features
8 enhancement help wanted

Robo-smart-Secure-System #38
Final Documentation
2 enhancement help wanted

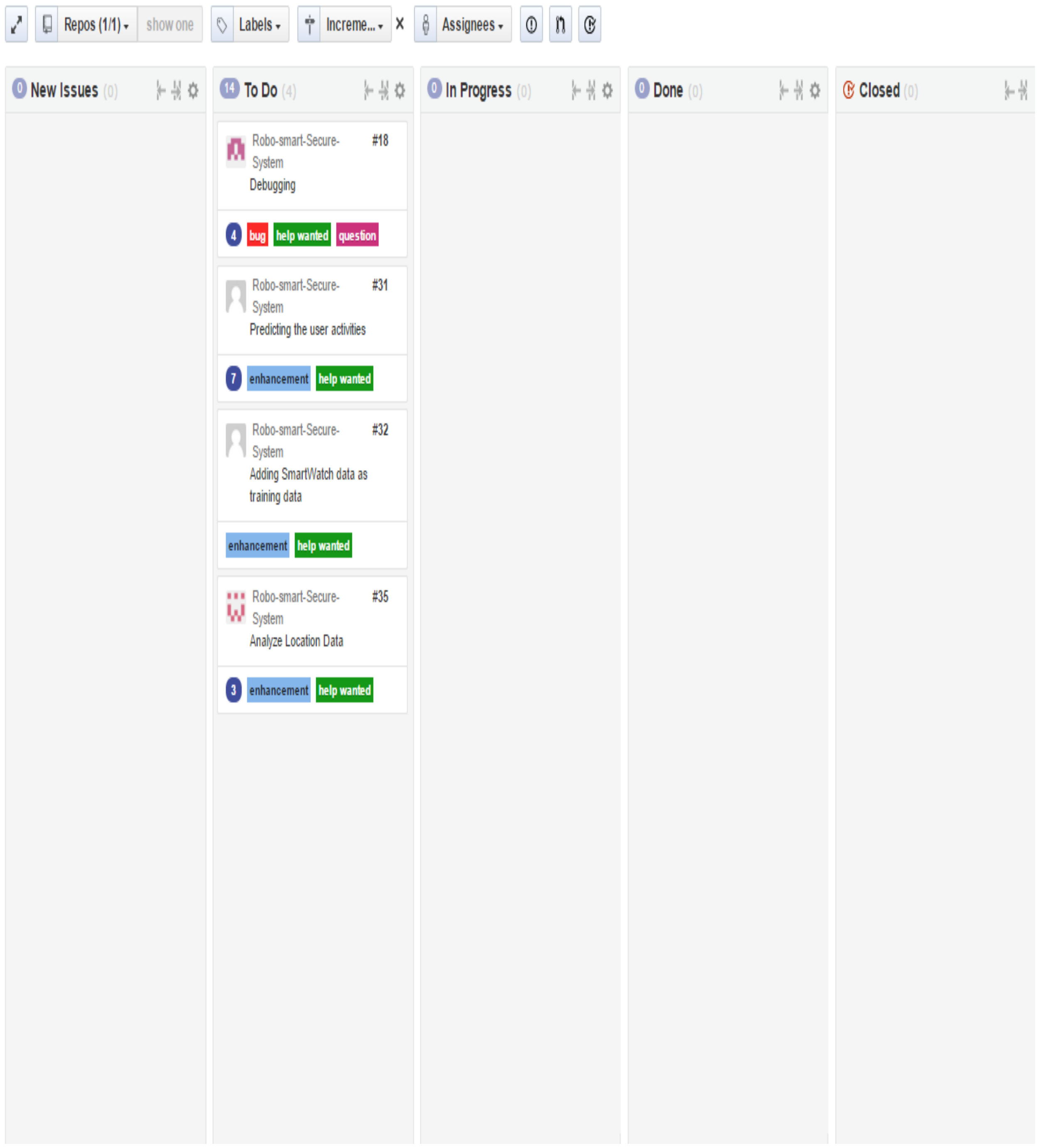
0 In Progress (0)

0 Done (0)

Closed (0)

Final Increment:

The timeline for the final increment due on 04/29/2016 will have the graph as follows:



0 New Issues (0)

14 To Do (4)

Robo-smart-Secure-System #18
Debugging
4 bug help wanted question

Robo-smart-Secure-System #31
Predicting the user activities
7 enhancement help wanted

Robo-smart-Secure-System #32
Adding SmartWatch data as training data
enhancement help wanted

Robo-smart-Secure-System #35
Analyze Location Data
3 enhancement help wanted

0 In Progress (0)

0 Done (0)

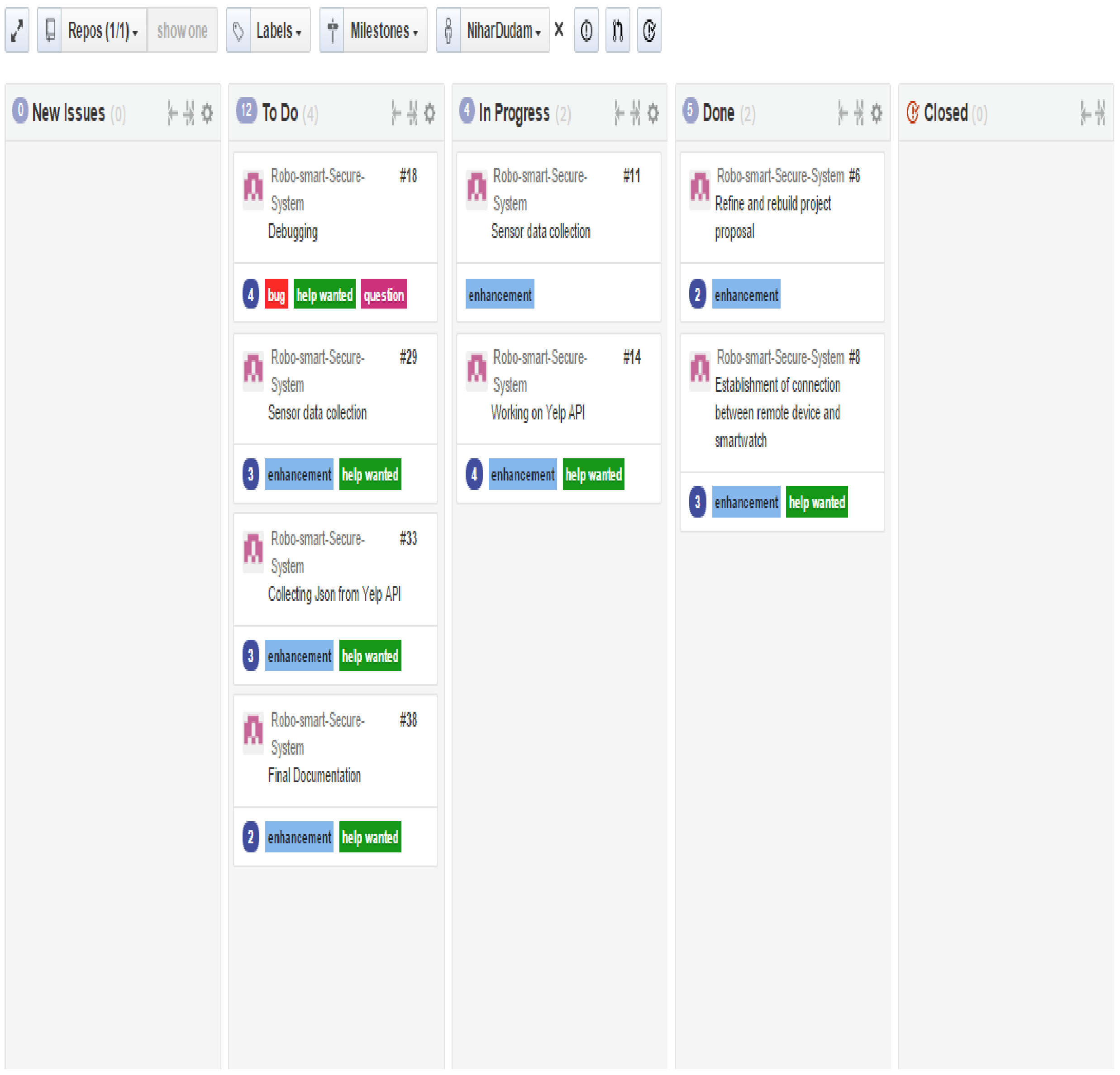
Closed (0)

3. Project Members and Task Responsibility:

This section deals with the timeline of the project for each member along with the task responsibility:

a. Nihar Dudam

The tasks assigned to Nihar in order to facilitate the completion of the project has the following diagram:



0 New Issues (0)

12 To Do (4)

Robo-smart-Secure-System #18
Debugging
4 bug help wanted question

Robo-smart-Secure-System #29
Sensor data collection
3 enhancement help wanted

Robo-smart-Secure-System #33
Collecting Json from Yelp API
3 enhancement help wanted

Robo-smart-Secure-System #38
Final Documentation
2 enhancement help wanted

4 In Progress (2)

Robo-smart-Secure-System #11
Sensor data collection
enhancement

Robo-smart-Secure-System #14
Working on Yelp API
4 enhancement help wanted

5 Done (2)

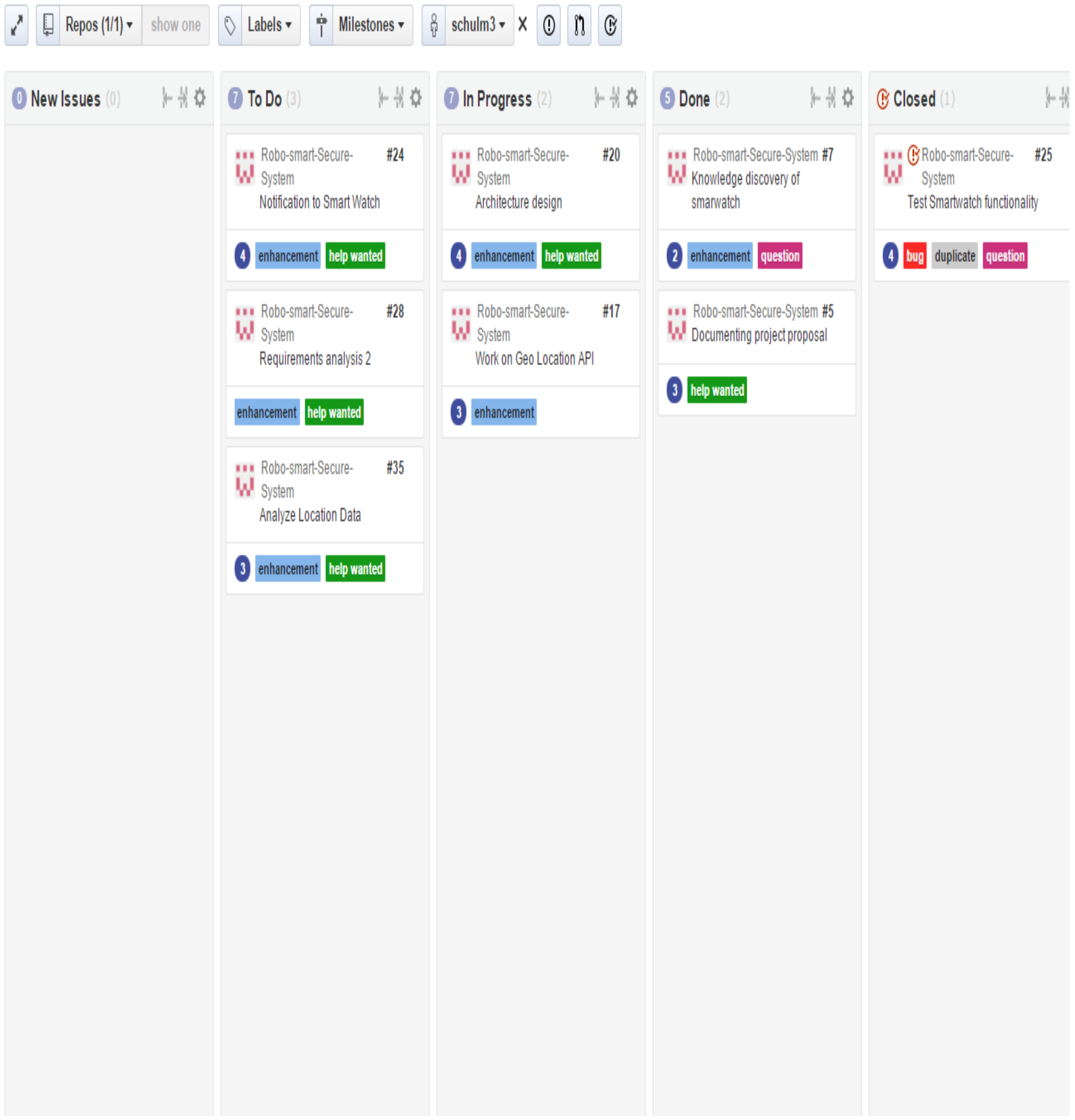
Robo-smart-Secure-System #6
Refine and rebuild project proposal
2 enhancement

Robo-smart-Secure-System #8
Establishment of connection between remote device and smartwatch
3 enhancement help wanted

Closed (0)

b. Mark Schultz

The tasks assigned to Mark in order to facilitate the completion of the project has the following diagram:


Repos (1/1) show one Labels Milestones schulm3 X

0 New Issues (0)

7 To Do (3)

Robo-smart-Secure-System #24
Notification to Smart Watch
4 enhancement help wanted

Robo-smart-Secure-System #28
Requirements analysis 2
enhancement help wanted

Robo-smart-Secure-System #35
Analyze Location Data
3 enhancement help wanted

7 In Progress (2)

Robo-smart-Secure-System #20
Architecture design
4 enhancement help wanted

Robo-smart-Secure-System #17
Work on Geo Location API
3 enhancement

5 Done (2)

Robo-smart-Secure-System #7
Knowledge discovery of smarwatch
2 enhancement question

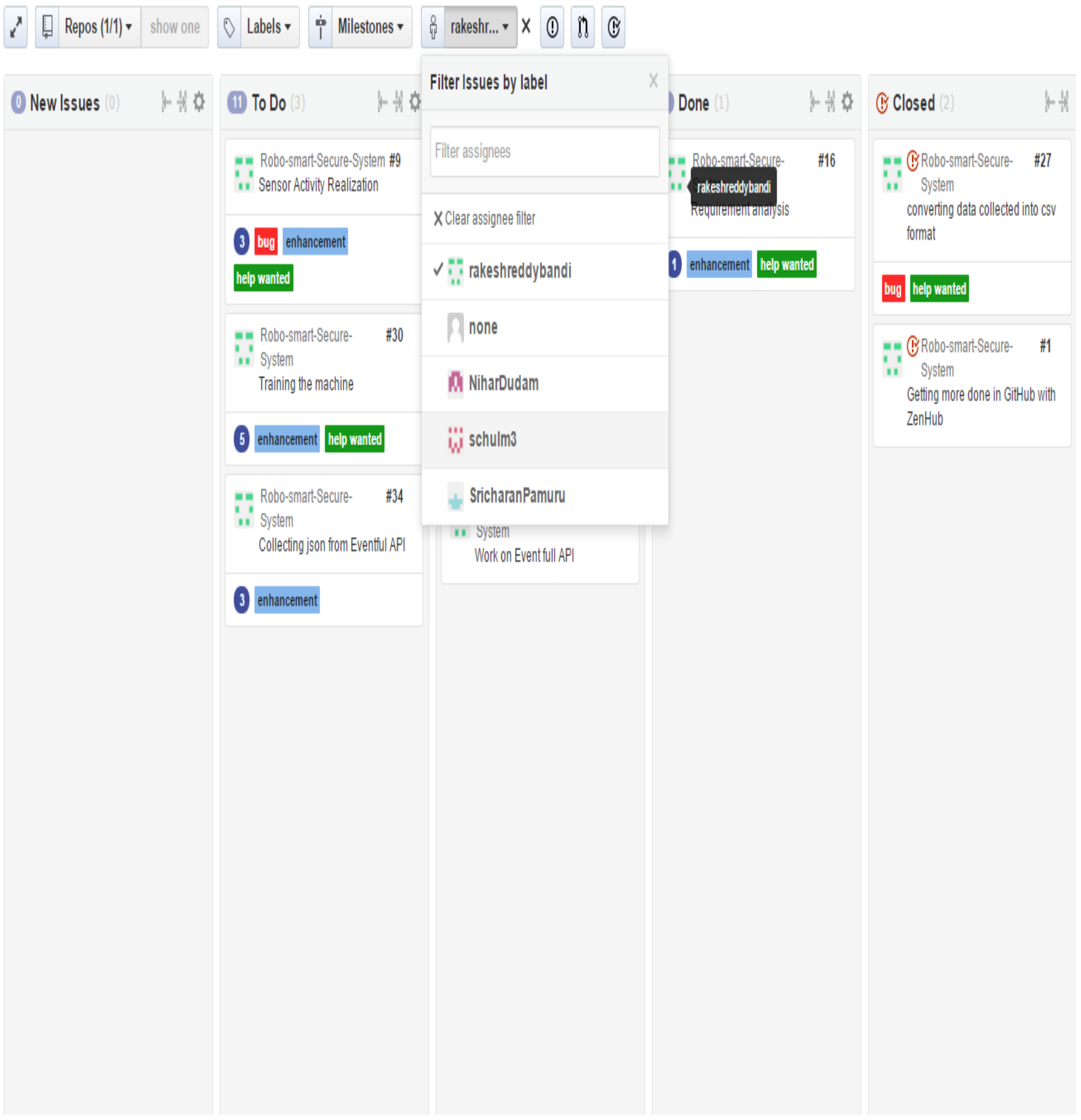
Robo-smart-Secure-System #5
Documenting project proposal
3 help wanted

Closed (1)

Robo-smart-Secure-System #25
Test Smartwatch functionality
4 bug duplicate question

c. Rakesh Reddy Bandi

The tasks assigned to Rakesh in order to facilitate the completion of the project has the following diagram:



Robo-smart-Secure-System #9
Sensor Activity Realization
3 bug enhancement help wanted

Robo-smart-Secure-System #30
Training the machine
5 enhancement help wanted

Robo-smart-Secure-System #34
Collecting json from Eventful API
3 enhancement

System
Work on Event full API

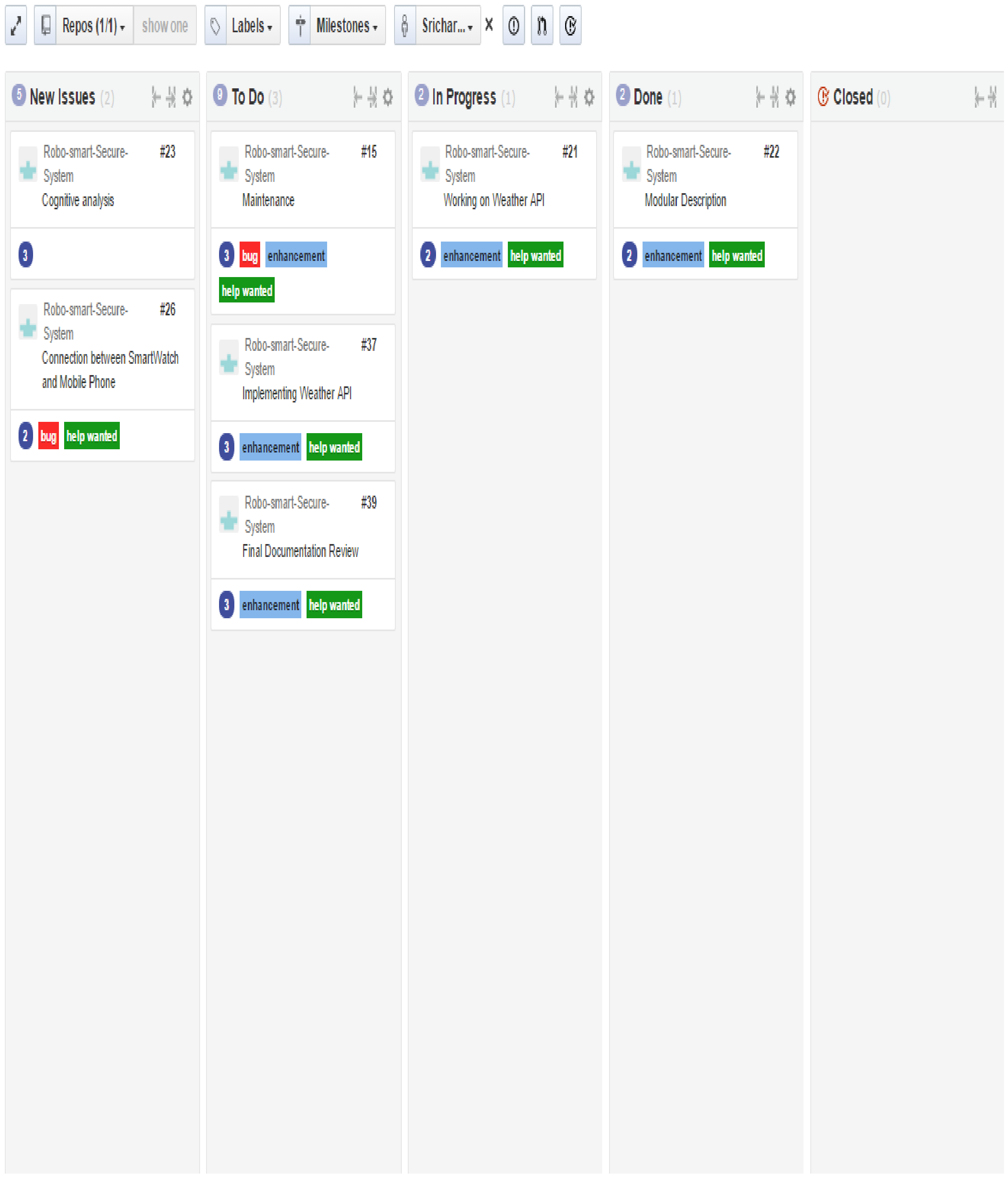
Robo-smart-Secure- #16
rakeshreddybandi
Requirement analysis
1 enhancement help wanted

Robo-smart-Secure-System #27
converting data collected into csv format
bug help wanted

Robo-smart-Secure-System #1
Getting more done in GitHub with ZenHub

d. Sricharan Pamuru

The tasks assigned to Sricharan in order to facilitate the completion of the project has the following diagram:



5 New Issues (2)

Robo-smart-Secure-System #23
Cognitive analysis
3

Robo-smart-Secure-System #26
Connection between SmartWatch and Mobile Phone
2 bug help wanted

0 To Do (3)

Robo-smart-Secure-System #15
Maintenance
3 bug enhancement help wanted

Robo-smart-Secure-System #37
Implementing Weather API
3 enhancement help wanted

Robo-smart-Secure-System #39
Final Documentation Review
3 enhancement help wanted

2 In Progress (1)

Robo-smart-Secure-System #21
Working on Weather API
2 enhancement help wanted

2 Done (1)

Robo-smart-Secure-System #22
Modular Description
2 enhancement help wanted

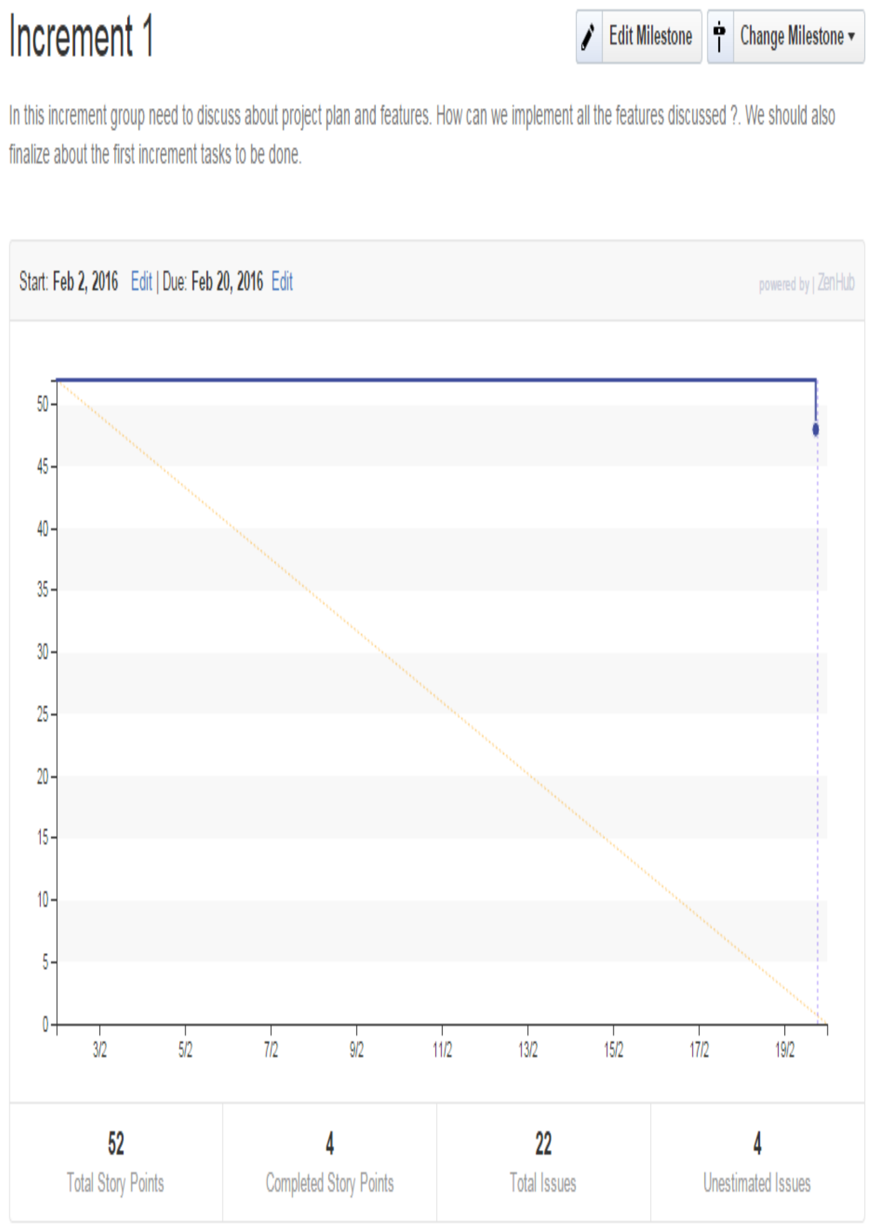
Closed (0)

4. Burndown charts:

The burndown charts would basically represent the time deployment of project as per the various phases. It would instantiate the modules of time and its according development over the schedule.

Increment 1:

The burndown chart of increment 1:



Increment 1

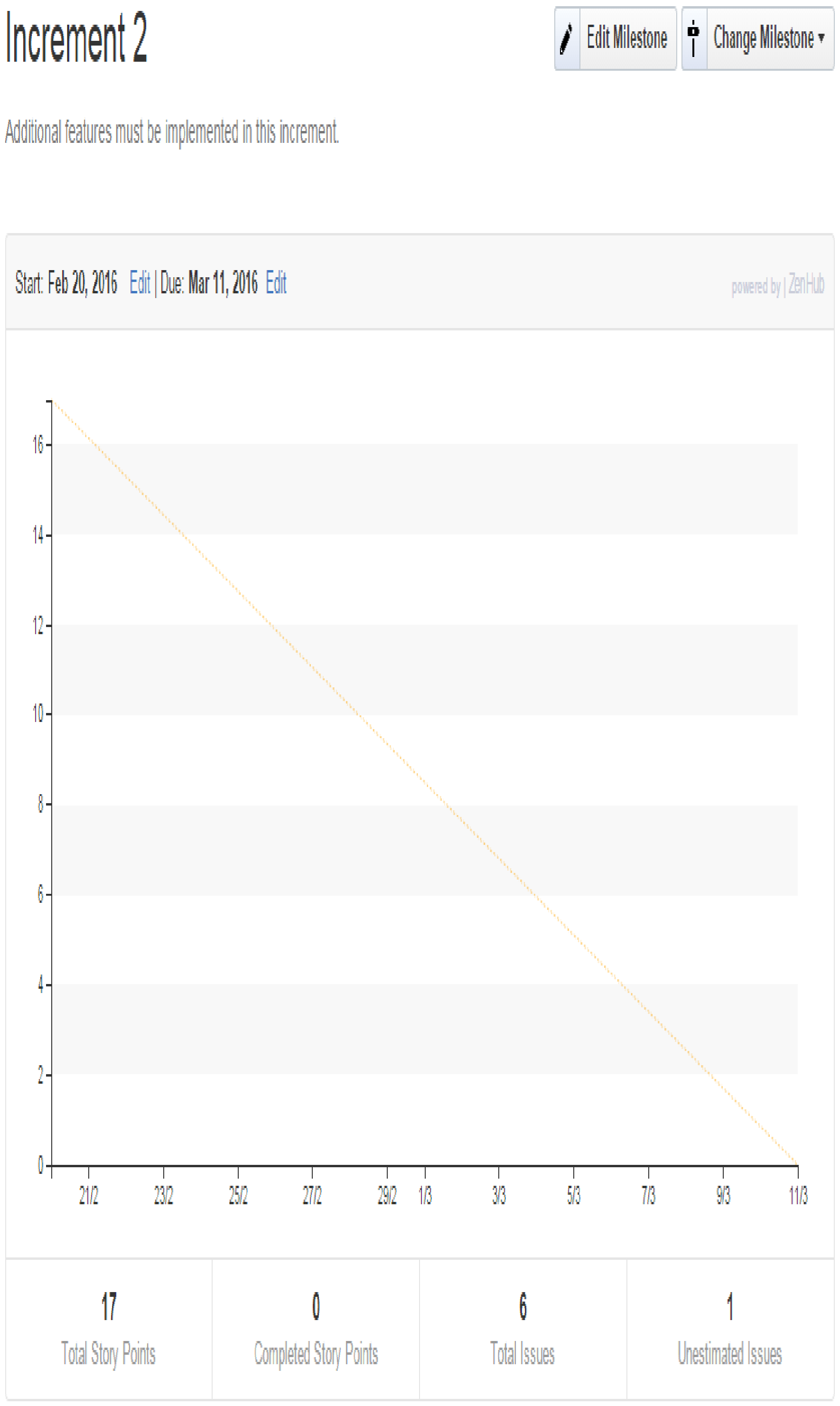
In this increment group need to discuss about project plan and features. How can we implement all the features discussed ?. We should also finalize about the first increment tasks to be done.

Start: Feb 2, 2016 Edit | Due: Feb 20, 2016 Edit

52	4	22	4
Total Story Points	Completed Story Points	Total Issues	Unestimated Issues

Increment 2:

The burndown chart of increment 2 is as follows:



Increment 2

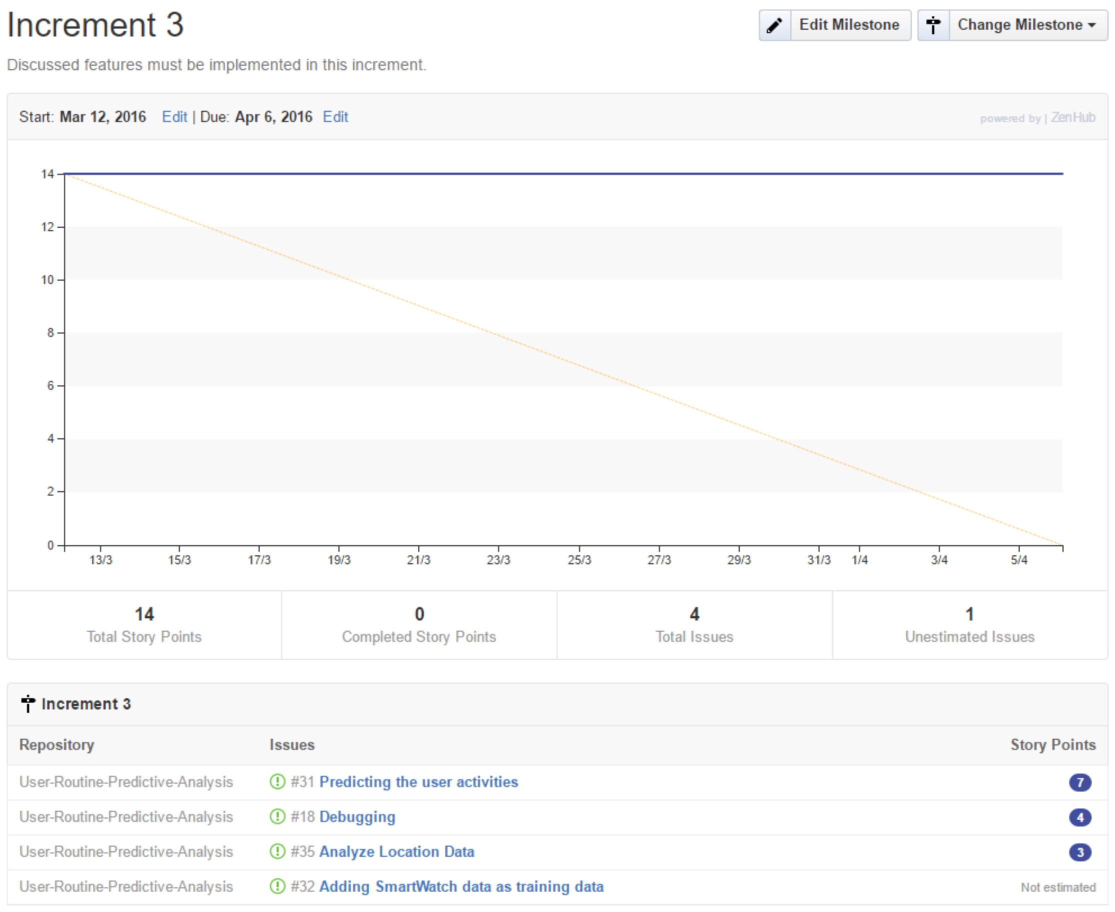
Additional features must be implemented in this increment.

Start: Feb 20, 2016 Edit | Due: Mar 11, 2016 Edit

17	0	6	1
Total Story Points	Completed Story Points	Total Issues	Unestimated Issues

Increment 3:

The burndown chart of increment 3 is:



Increment 3

Discussed features must be implemented in this increment.

Start: **Mar 12, 2016** Edit | Due: **Apr 6, 2016** Edit

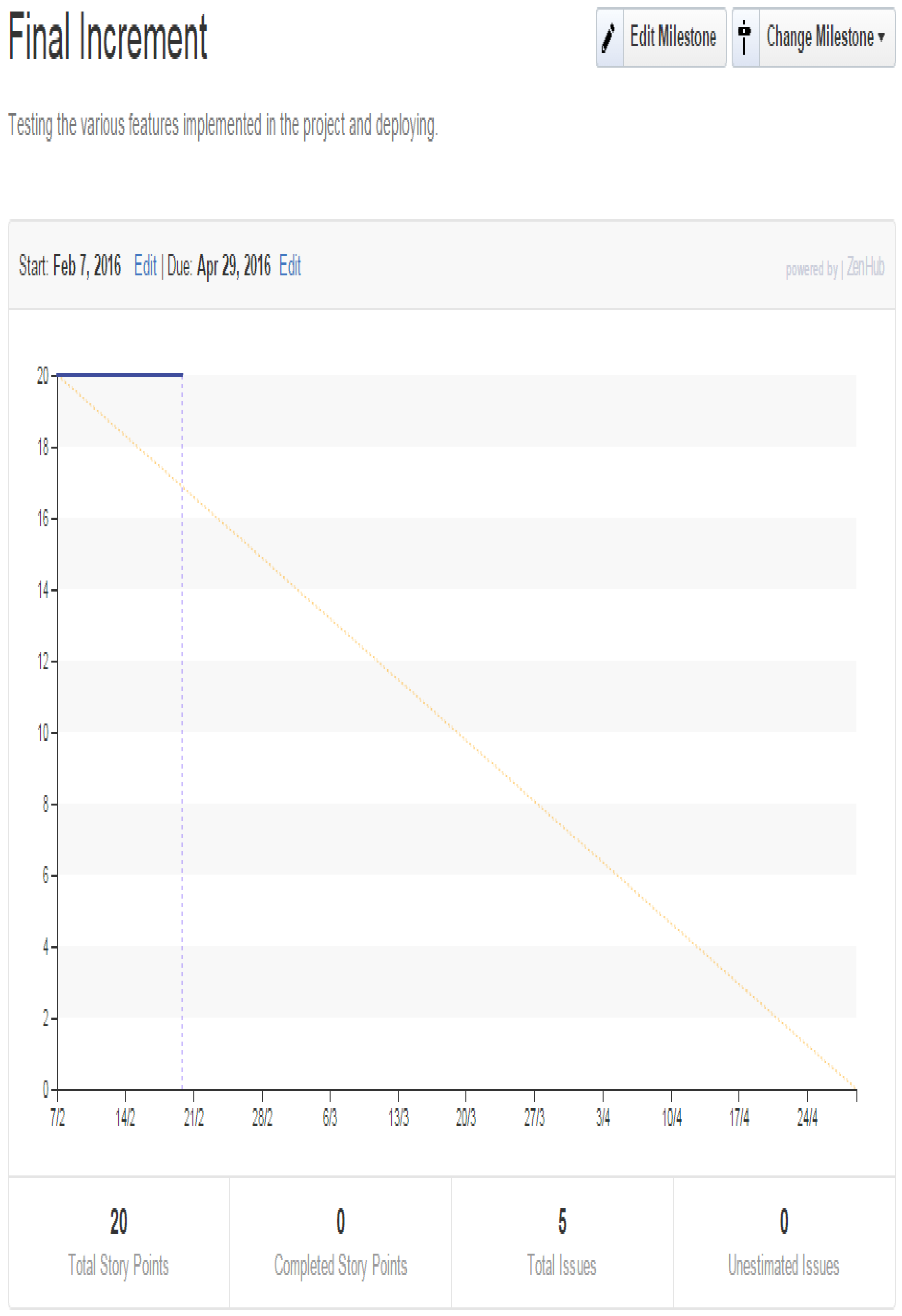
14	0	4	1
Total Story Points	Completed Story Points	Total Issues	Unestimated Issues

Increment 3

Repository	Issues	Story Points
User-Routine-Predictive-Analysis	#31 **Predicting the user activities**	7
User-Routine-Predictive-Analysis	#18 **Debugging**	4
User-Routine-Predictive-Analysis	#35 **Analyze Location Data**	3
User-Routine-Predictive-Analysis	#32 **Adding SmartWatch data as training data**	Not estimated

Final Increment:

The final increment has the burndown chart as follows:


Final Increment

Testing the various features implemented in the project and deploying.

Start: Feb 7, 2016 Edit | Due: Apr 29, 2016 Edit

20	0	5	0
Total Story Points	Completed Story Points	Total Issues	Unestimated Issues

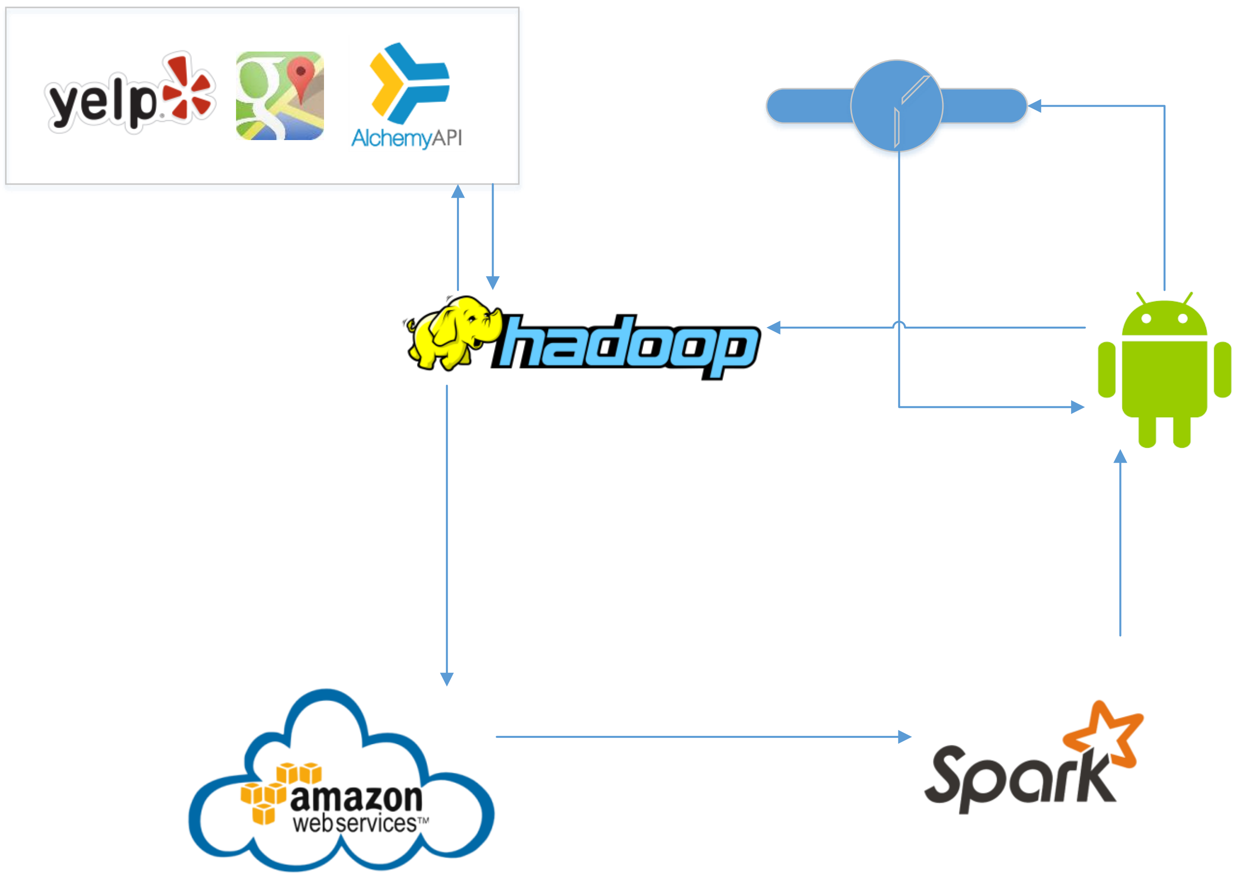
IV. Third Increment Report:

In this increment, we have implemented several pieces of our project. First, we have successfully used k-means clustering to establish frequent locations of our user, given information about latitude and longitude. We have tested this data by creating a large set of input data around a specific point in downtown Kansas City. As expected, the clusters are evenly dispersed around said point. When inserting one specific location into the data multiple times, the clustering method adjusts accordingly.

Once the cluster centers are found, the clusters then receive labels. The clusters detect frequent locations, and the labels provide additional value to the said location. We export the k centers to Google Maps API to determine the address of each center. Furthermore, we use this to determine the type of location (bookstore, library, coffee shop, etc.).

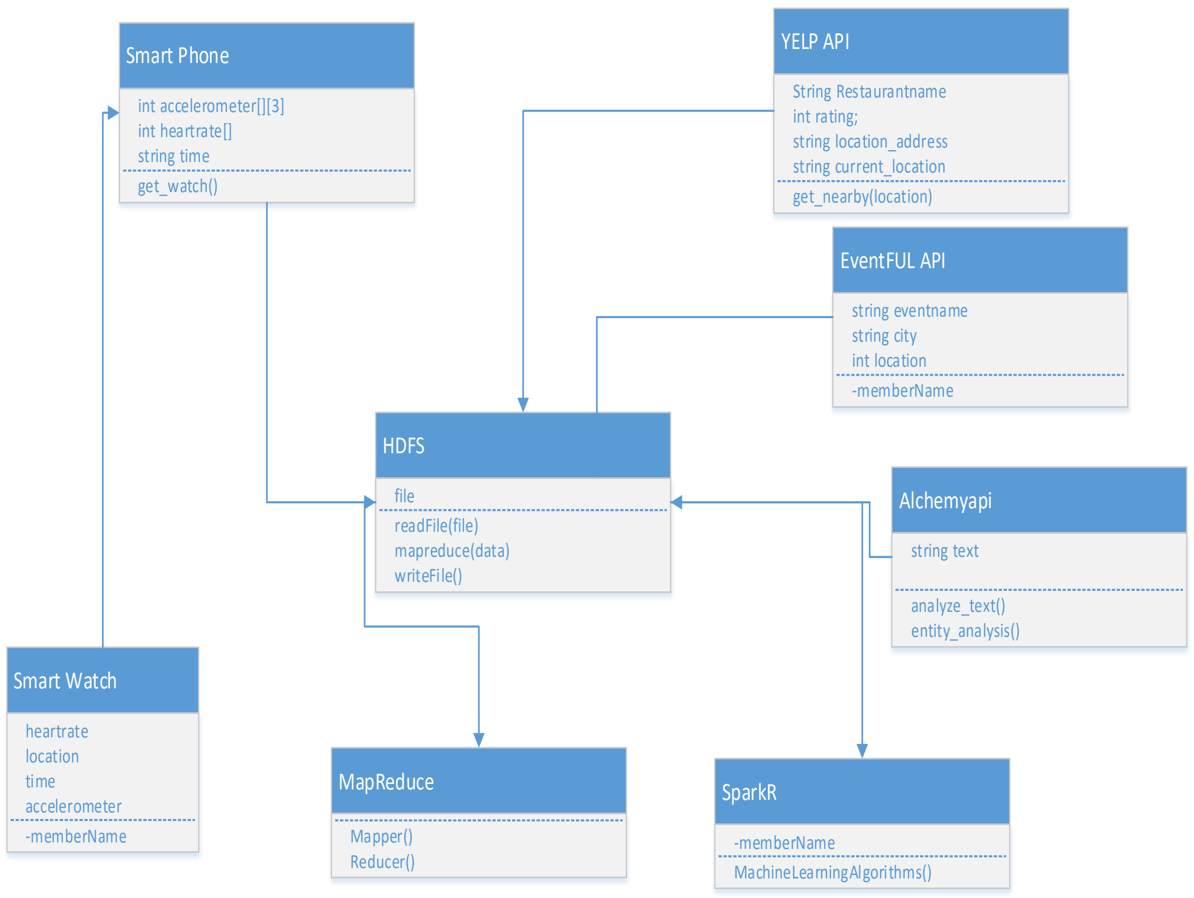
Also, we have built a socket connection which live streams data from the mobile app to the spark stream. With the spark stream, we are able to implement our naïve Bayesian model in order to classify specific movements based on accelerometer data. Such movements include walking, eating, and climbing the stairs. While these small movements may not seem like provide much insight alone, we will be using it as data to predict even larger trends in routines. The socket also provides other data, including the latitudes and longitudes for location detection, time, and day.

1. **Design of Features:**
    a. **Architecture:**



Our application begins and ends with the smartwatch user interface. The user will input to the watch. watch sends data to the Hadoop File System through the Android application. The Hadoop System will organize data that comes both from the user and from APIs. The application will then send data to the Amazon Web Services to run analytics on the data. We will implement MapReduce programs in Spark and Spark R. We will use several machine learning algorithms for prediction. After running the Query in spark, the data is sent back to the Android application, which can then be viewed by the user on their smartwatch.

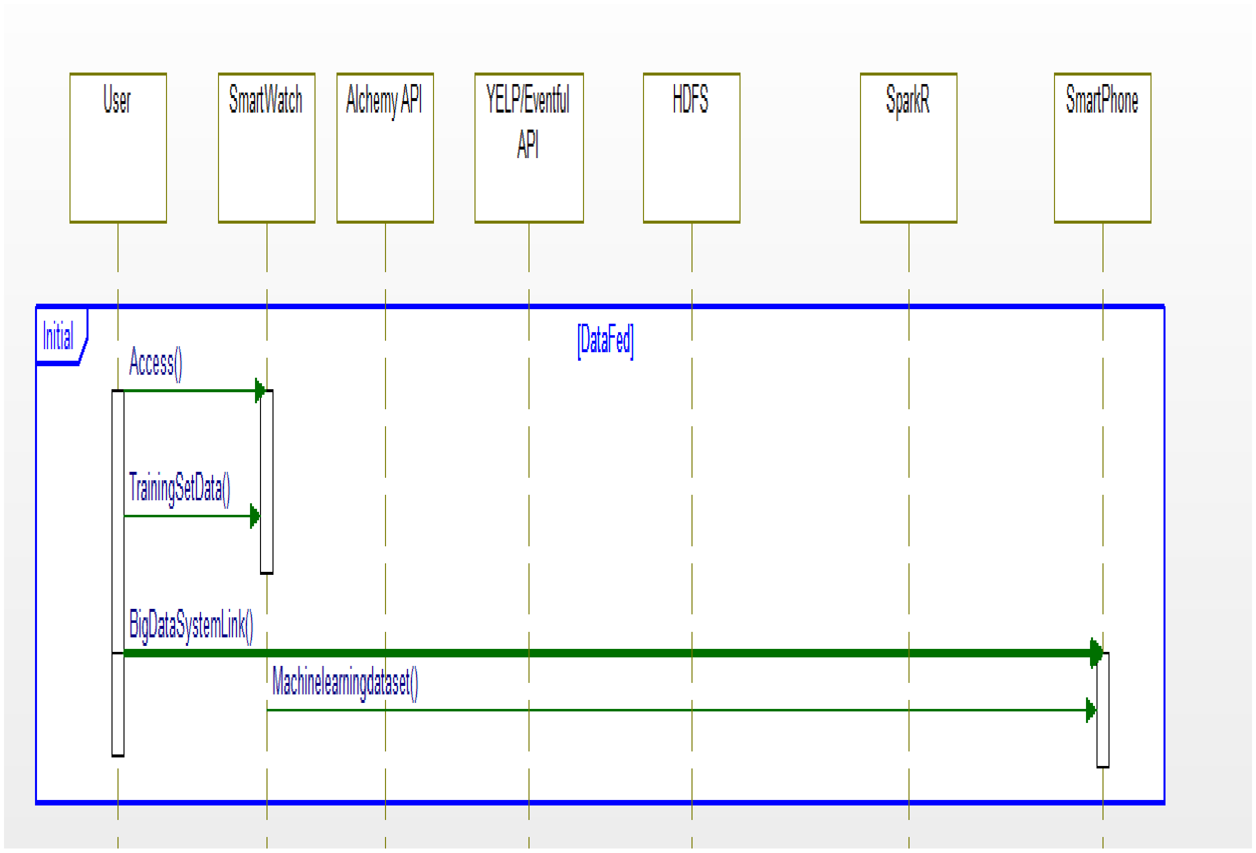
b. Class Diagram:



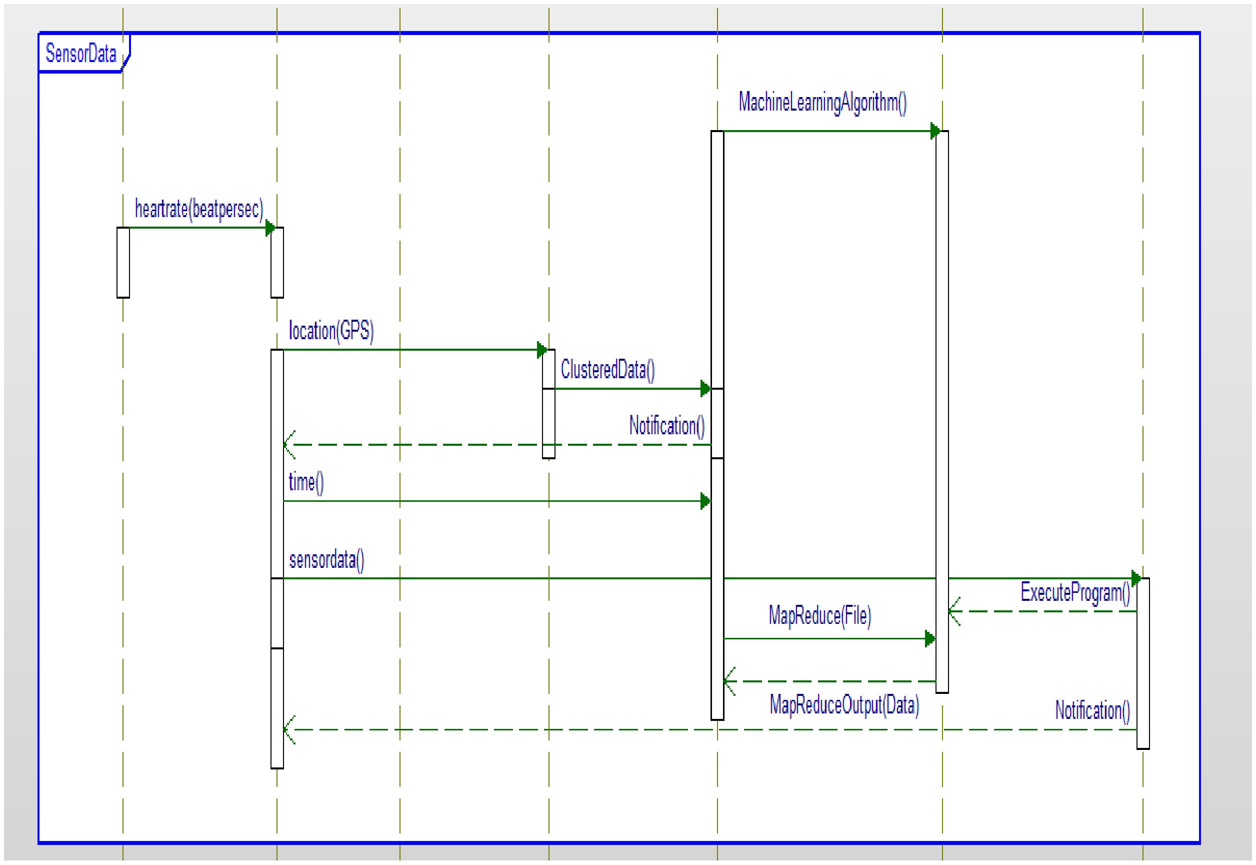
The class diagram we have drawn outlines the structure of our current project. We have three classes for three APIs, which provide data to the HDFS class. Also providing data to the HDFS is the smart phone, which calls get_watch() to retreive smart watch data. The smart watch collects heartrate, location, time, and accelerometer data. HDFS will implement MapReduce and SparkR classes to analyze data it is given. SparkR will implement several of the machine learning algorithms, which will be discussed in more detail later in this paper.

c. Sequence Diagram:

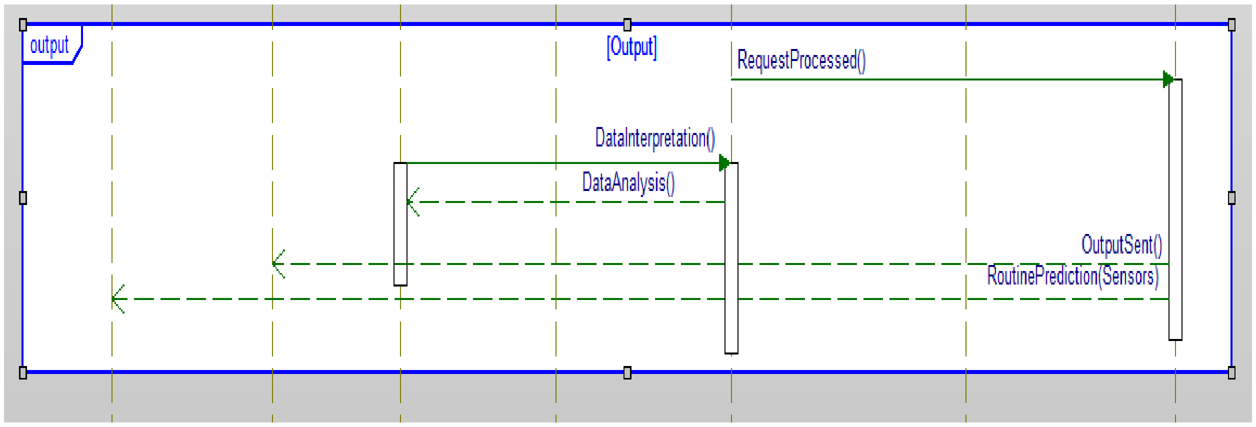
The sequence diagram is shown as below. The sequence diagram is Disintegrated and attached:



The datafed process would mainly consist of all the classes to initially set up communication. The access and training set data is collected and then related with the smartwatch by the user as there is a setup of data fed into the remote machine. The machine learning dataset would be communicated through the smart phone and the device learns as to what the user pattern is.



The sensor data would basically do the operations involved mainly from the HDFS to obtain a programmed output that provides the user as to what the routine is or rather should be.



The final output would basically be a clear cut approach and the user gets a routine predicted.

d. **Machine Learning Algorithms:**

In Increment one, we have implemented a naïve bayes algorithm in R to classify datasets of user movements. The naïve bayes is a classifier that uses Bayes theorem to determine the most likely outcome. Bayes Theorem is written as such: $P(A|B) = \frac{P(A)*P(B|A)}{P(B)}$. In this equation, P(A) is the prior, P(B|A) is the likelihood, and P(B) is the evidence. This algorithm relies on P(A) and P(B) being independent of one another. While in reality, this is rarely the case, it provides an estimate that is extremely effective in classification.

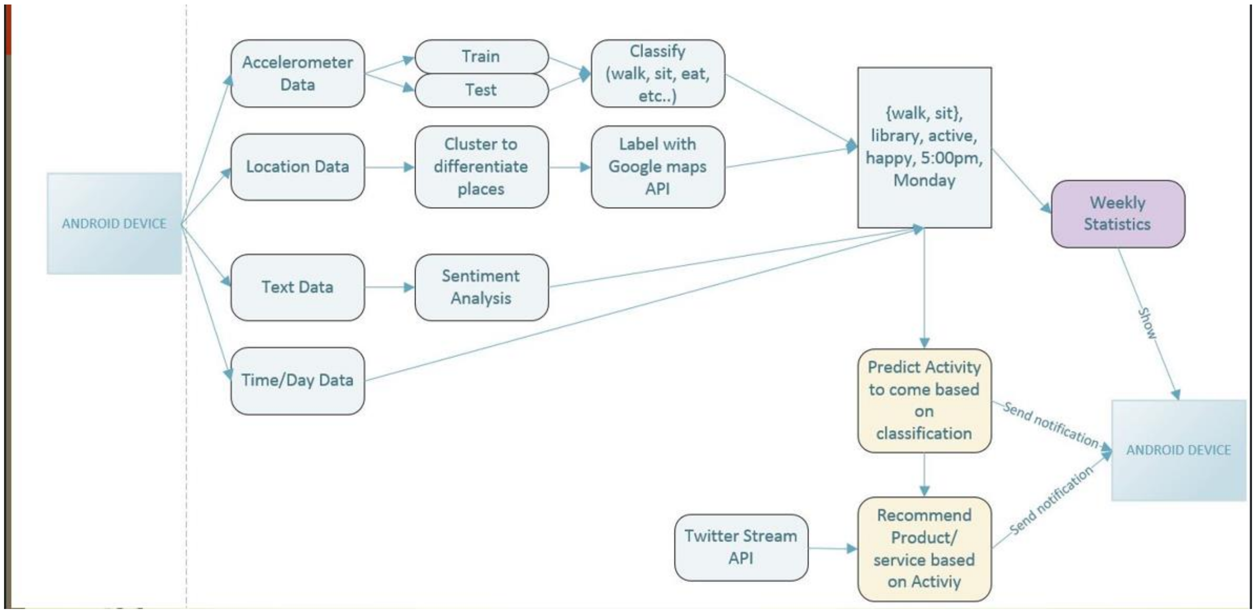
In future increments, we plan on using a several more machine learning algorithms implemented on SparkR to scale larger sets of data.

e. **Datasets:**

A dataset is a collection of related sets of information iterated across an activity centric output. The dataset collected as input from the smartwatch will be fed via machine learning algorithms to the remote machine for it to be analyzed. The data/information is obtained from the different sensors which are required for providing an analytic comparison. Output is retrieved from the values fed. Supervised learning is in the initial stages since the machine is trained or rather intelligence provided regarding what to react in which case. After the training data set is provided, unsupervised learning comes to the frame. Here, the machine is able to analyze the data and compare with the training set. It will also be able to learn in case the user does an activity different from the routine. A notification is sent but nonetheless learning is also done by the machine, which in our case is the smart phone connected via USB.

2. Implementation

Workflow of project:


ANDROID DEVICE
Accelerometer Data
Location Data
Text Data
Time/Day Data
Train
Test
Cluster to differentiate places
Sentiment Analysis
Classify (walk, sit, eat, etc..)
Label with Google maps API
{walk, sit}, library, active, happy, 5:00pm, Monday
Weekly Statistics
Show
Predict Activity to come based on classification
Send notification
ANDROID DEVICE
Twitter Stream API
Recommend Product/ service based on Activiy
Send notification

Mobile Client Implementation (Smart Watch, Smart Phone, Robot):

Mobile Application for Collecting the Sensor Data:

- We have developed an application which is specifically for collecting the user smart phone and smart watch sensor data.
- In this application we have collected the Sensor streaming data and sending to Spark.
- Later we use this data to train over model to predict the user patterns regularly.
- This application mainly gets the all sensor data values from the Sensors such as Gyroscope, Accelerometer, Proximity, Light, Magnetic and Heart Rate in the Smart watch and smart phone.
- Apart from this we also collected the Location's latitude and longitude at which the sensor data Collected is sent to Spark.

Clustering the location data using K-means clustering technique:

- We have used k-means clustering technique to estimate user's frequent locations.
- Tested this clusters by creating large input data around KC-downtown.
- Once cluster centers are found they are labeled according to time zone.
- We export k centers to Google Maps API to determine the address of each center which can be used to determine the type of location.

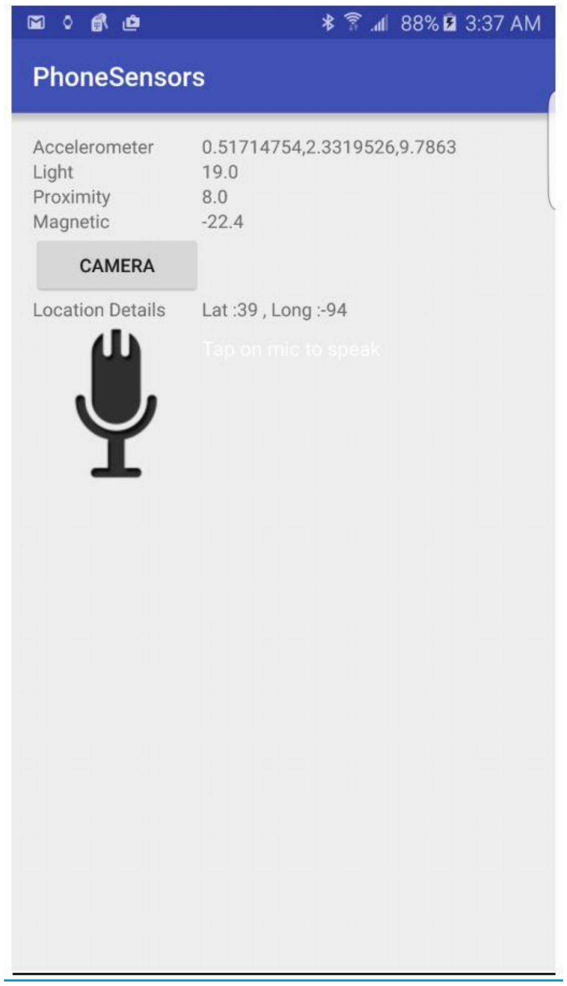
Machine Learning Application:

Machine Learning Classification Algorithm- Naïve Bayes Classification Algorithm to predict the User Activity:

- We have predicted the user activity using the Naïve Bayes Classification algorithm.
- We have trained the model based on the data obtained for different activities such as walking, getting out of bed and climbing up the stairs.
- So using the training data, we have predicted the user activity for the particular accelerometer readings using the Naïve Bayes Classification algorithm.
- Naïve Bayes Classification algorithm:
    1. Calculate the Mean and Standard deviation for each of column in the data set.
    2. The likelihood for each of the axis for activity is calculated.
    3. The likelihood of each of the axis reading was calculated using the Gaussian distribution function.
    4. Using the Naïve Bayes posterior probability formula, we have computed the probability of occurrence of posterior.
    5. Posterior probability is directly proportional to product of prior probability and likelihood.
- Using the posterior probability, we have predicted which activity is most likely to perform.
- The Machine learning algorithm runs using Spark

4. Deployment:

1. SmartPhone/SmartWatch Sensors Data Storage Application:



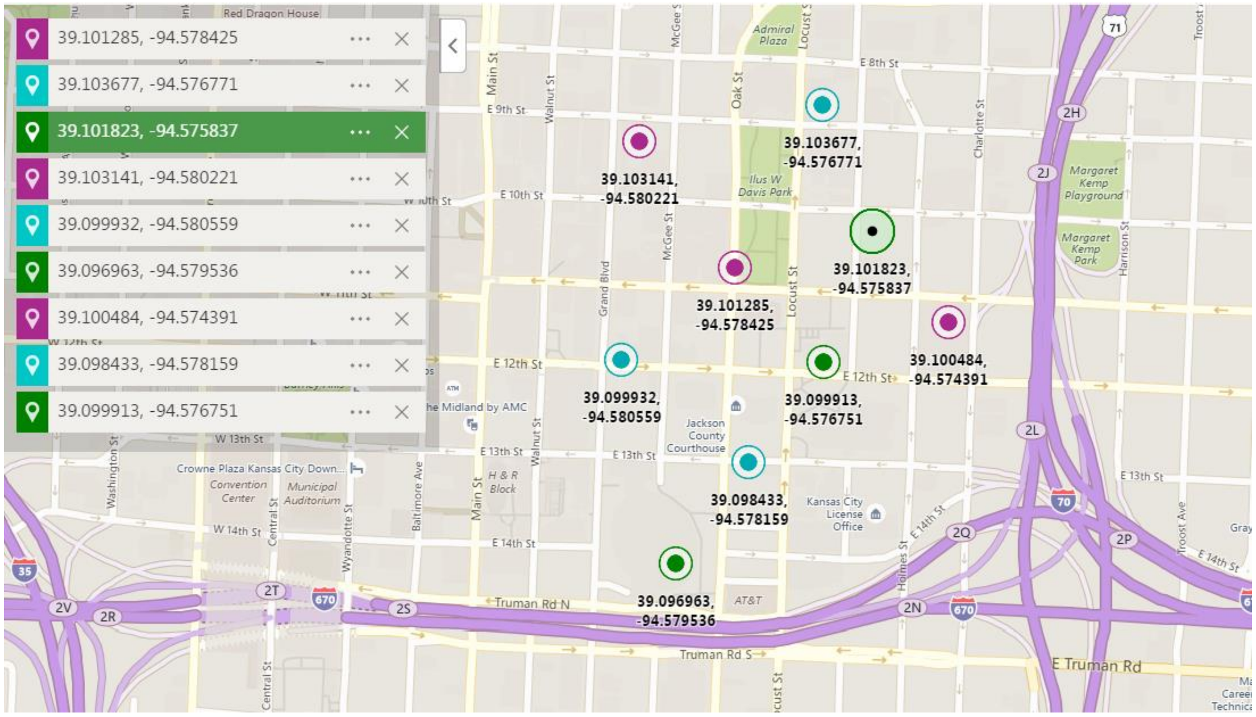
Accelerometer 0.51714754,2.3319526,9.7863
Light 19.0
Proximity 8.0
Magnetic -22.4

Location Details Lat :39 , Long :-94

The various sensors in both mobile and smart watch gives the data values continuously for the user activities.

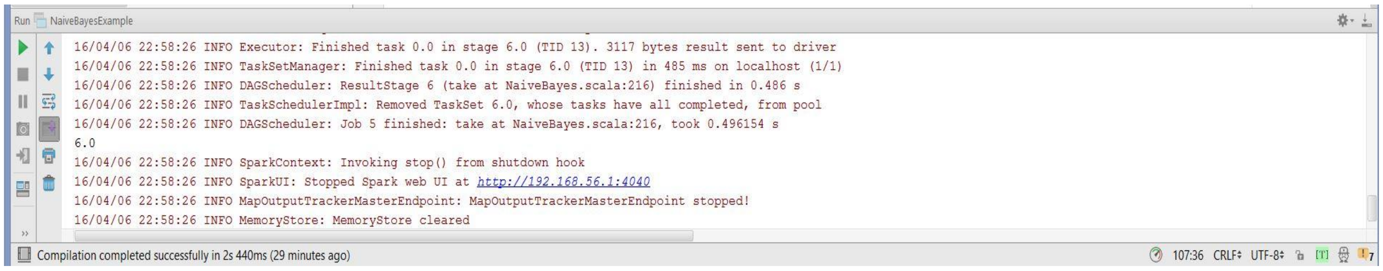
We store these values in the internal storage of the Device.

2. Clustering the location data using k-means:



Clusters created near KC-Downtown.

3. Displaying current activity of user activity in Spark where the data is taken from the Device



```
16/04/06 22:58:26 INFO Executor: Finished task 0.0 in stage 6.0 (TID 13). 3117 bytes result sent to driver
16/04/06 22:58:26 INFO TaskSetManager: Finished task 0.0 in stage 6.0 (TID 13) in 485 ms on localhost (1/1)
16/04/06 22:58:26 INFO DAGScheduler: ResultStage 6 (take at NaiveBayes.scala:216) finished in 0.486 s
16/04/06 22:58:26 INFO TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool
16/04/06 22:58:26 INFO DAGScheduler: Job 5 finished: take at NaiveBayes.scala:216, took 0.496154 s
6.0
16/04/06 22:58:26 INFO SparkContext: Invoking stop() from shutdown hook
16/04/06 22:58:26 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040
16/04/06 22:58:26 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/06 22:58:26 INFO MemoryStore: MemoryStore cleared
```

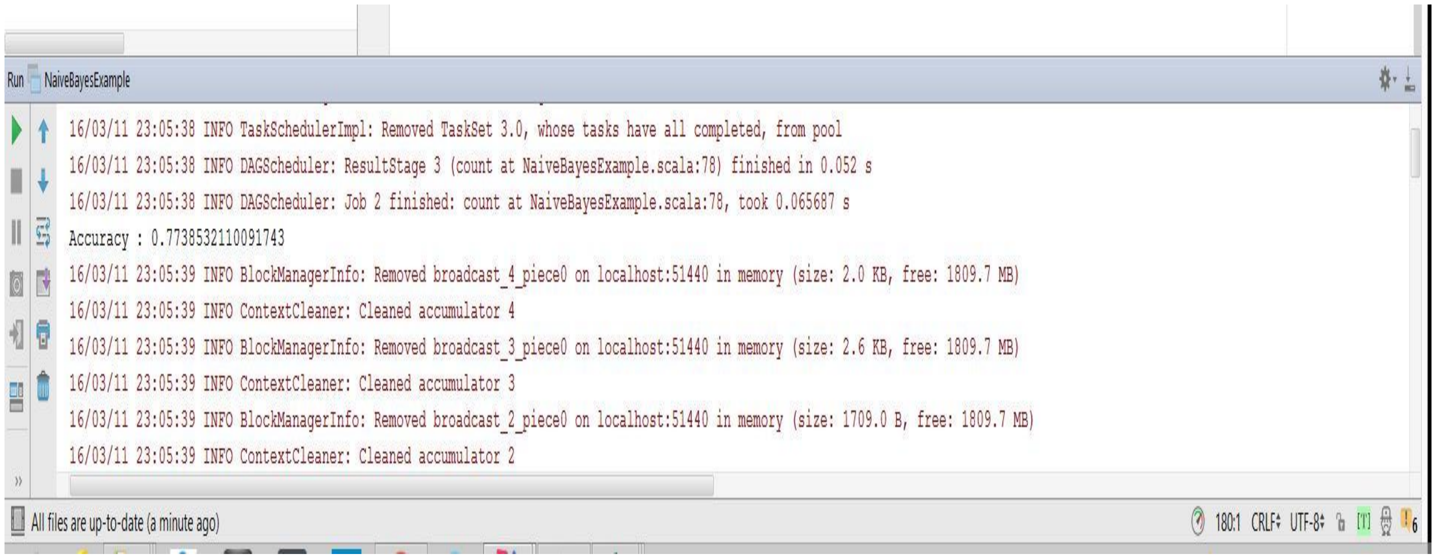
Compilation completed successfully in 2s 440ms (29 minutes ago)

Supervised Learning Algorithm Machine Learning Model:

Classification of User movements and activities using the Naïve Bayes Classification:

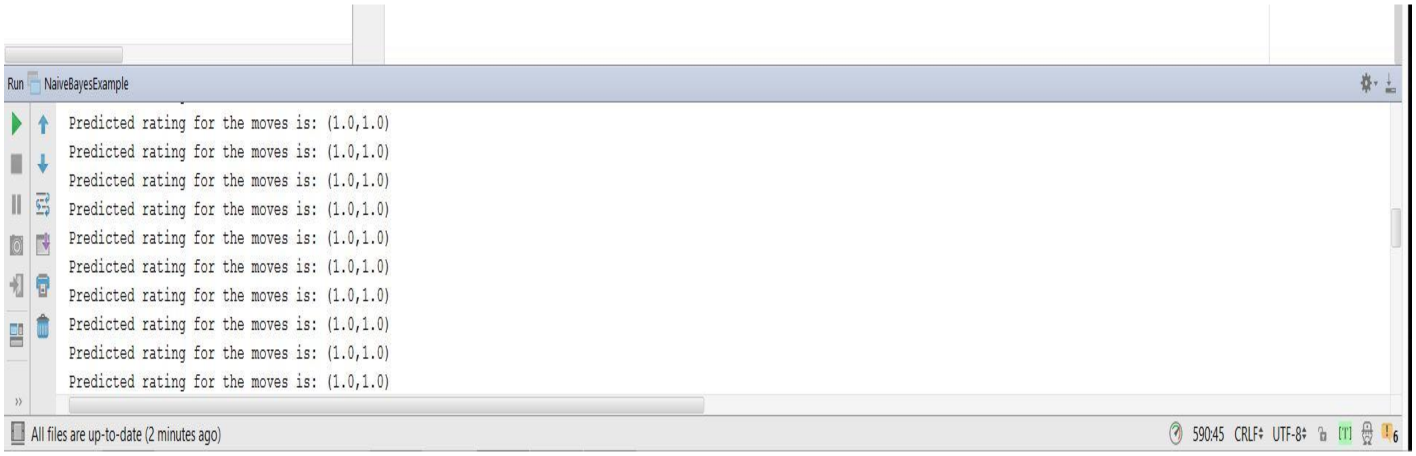
- We have divided the data into two parts testing data and training data.
- Training data—60%
- Testing data ---40%
- We have trained the model to predict the different user activities and movements such as sitting up the chair, sitting down the chair, walking, sleeping etc.,
- We used the Multinomial distribution and we have **normalized** the input data values to fall between 0-1 to increase the accuracy.
- Normalized is done using MLLib function.
- After Normalizing the features, We have converted the RDD into **labelled Vector RDD**
- Accuracy is calculated by validating the results predicted from the testing data to the actual values. We have got accuracy of 77.38%
- We can use this model to classify the incoming data from the user.
- We have saved this model to predict the later data from the user. From this data we provide the overall stats of user time spending on activities.

1. **Model Accuracy – nearly 77%:**



```
16/03/11 23:05:38 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/03/11 23:05:38 INFO DAGScheduler: ResultStage 3 (count at NaiveBayesExample.scala:78) finished in 0.052 s
16/03/11 23:05:38 INFO DAGScheduler: Job 2 finished: count at NaiveBayesExample.scala:78, took 0.065687 s
Accuracy : 0.7738532110091743
16/03/11 23:05:39 INFO BlockManagerInfo: Removed broadcast_4_piece0 on localhost:51440 in memory (size: 2.0 KB, free: 1809.7 MB)
16/03/11 23:05:39 INFO ContextCleaner: Cleaned accumulator 4
16/03/11 23:05:39 INFO BlockManagerInfo: Removed broadcast_3_piece0 on localhost:51440 in memory (size: 2.6 KB, free: 1809.7 MB)
16/03/11 23:05:39 INFO ContextCleaner: Cleaned accumulator 3
16/03/11 23:05:39 INFO BlockManagerInfo: Removed broadcast_2_piece0 on localhost:51440 in memory (size: 1709.0 B, free: 1809.7 MB)
16/03/11 23:05:39 INFO ContextCleaner: Cleaned accumulator 2
```

2. Displaying Confusion Matrix for the User Activity :



```
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
Predicted rating for the moves is: (1.0,1.0)
```

5. Project Management

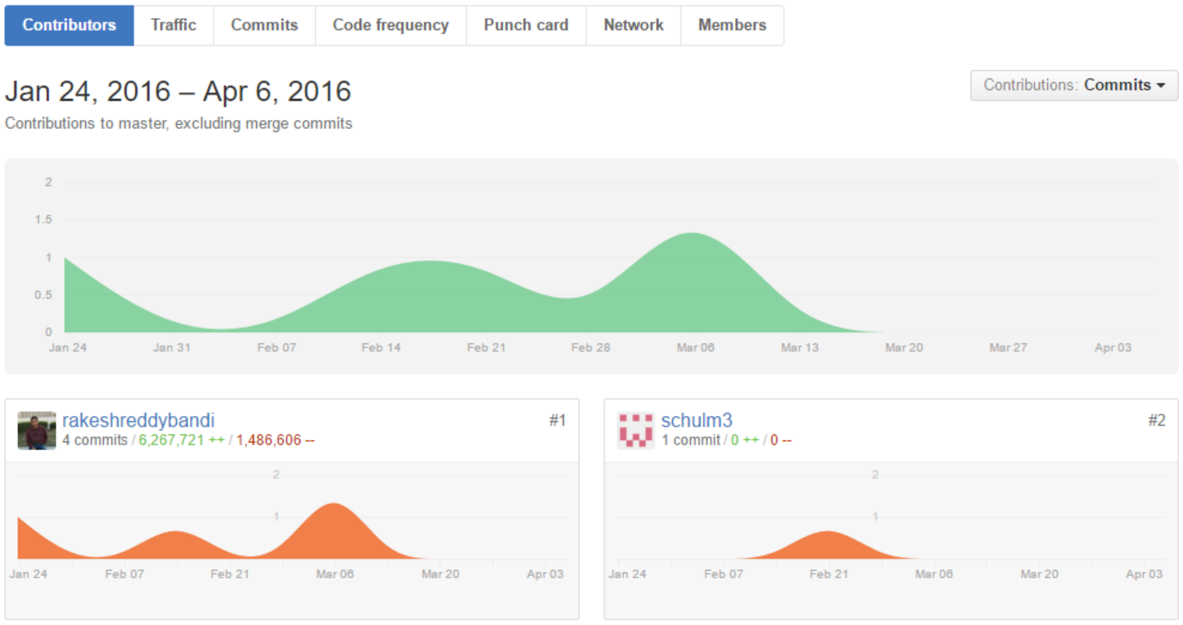
Increment 3:

The following image shows how the image is categorized and the implementation using the Zenhub Tool.

Graphs:

The following graphs show the analysis of the project management done over the period between first and second increment.

Contributors:



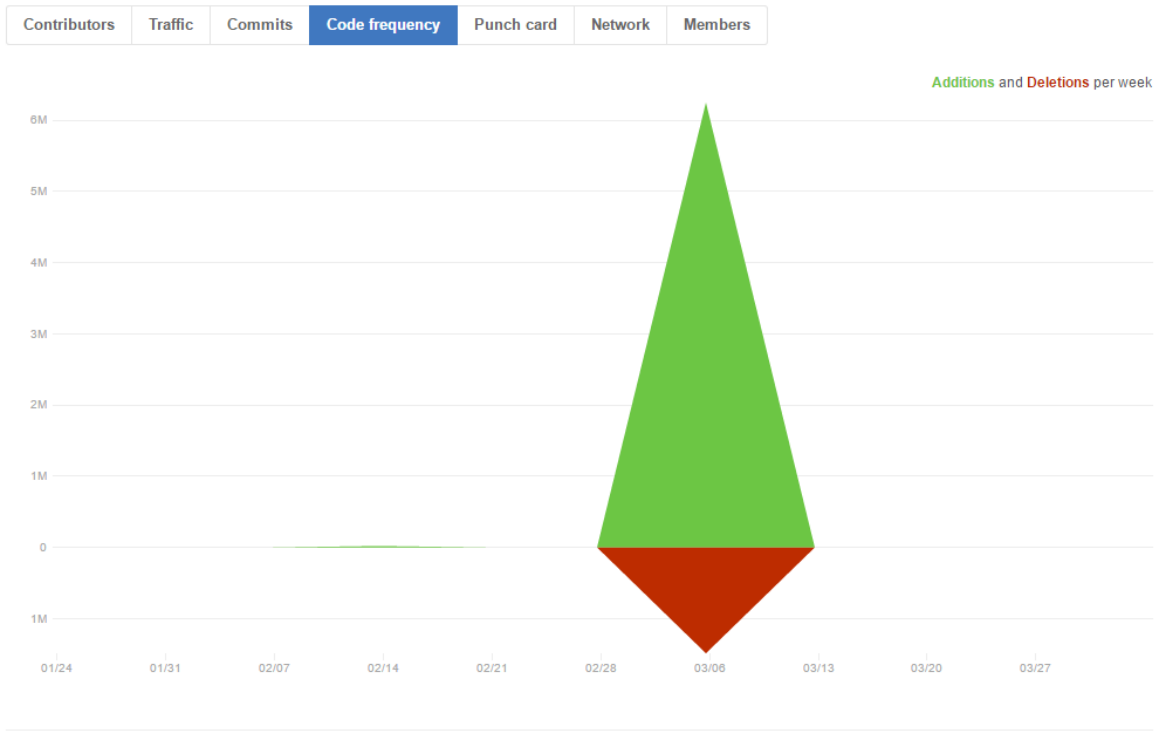
Jan 24, 2016 – Apr 6, 2016

Contributions to master, excluding merge commits

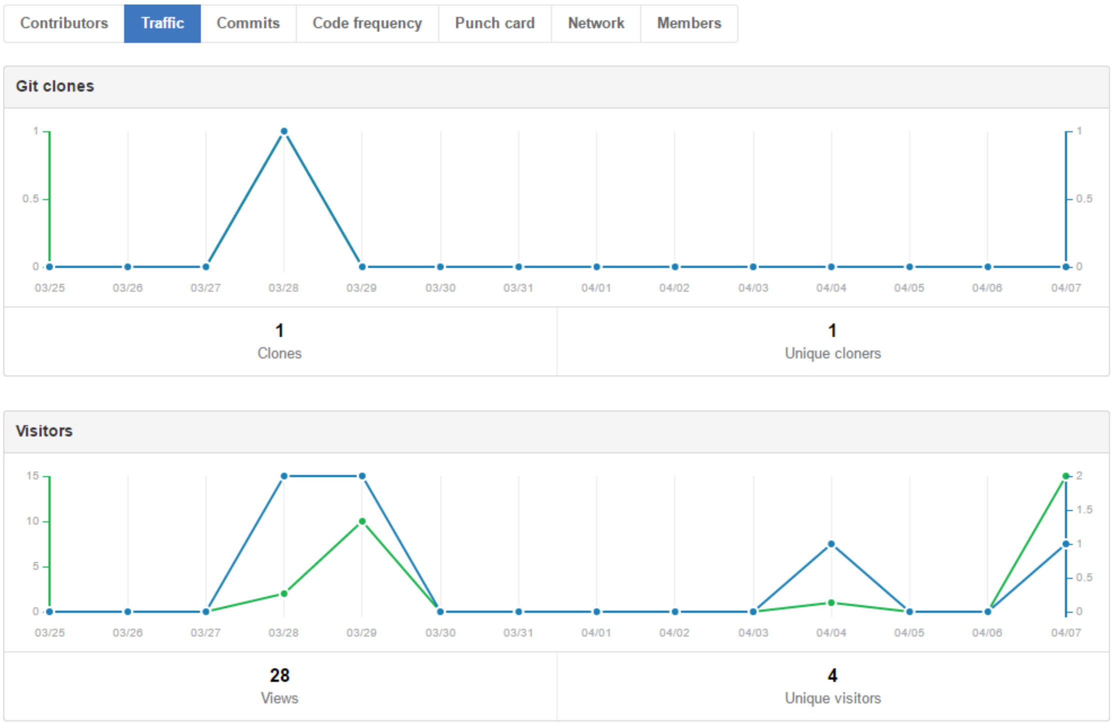
rakeshreddybandi
4 commits / 6,267,721 ++ / 1,486,606 --

schulm3
1 commit / 0 ++ / 0 --

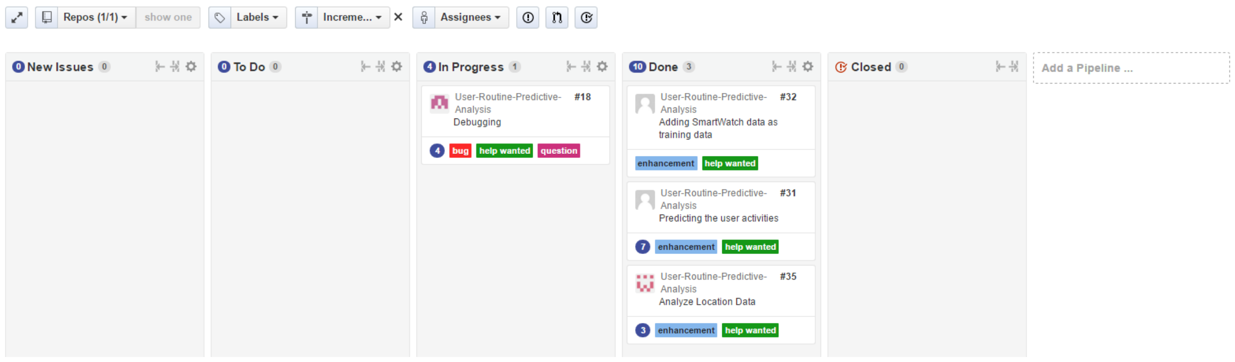
Code Frequency:



Traffic :



Increment3 ZenHub tool:



0 New Issues 0

0 To Do 0

4 In Progress 1

User-Routine-Predictive-Analysis #18
Debugging
4 bug help wanted question

10 Done 3

User-Routine-Predictive-Analysis #32
Adding SmartWatch data as training data
enhancement help wanted

User-Routine-Predictive-Analysis #31
Predicting the user activities
7 enhancement help wanted

User-Routine-Predictive-Analysis #35
Analyze Location Data
3 enhancement help wanted

Closed 0

- Implementation status report:

a. Work completed

Member	Task Description	Member Responsibility	Contribution %	Time (hours)	Comments /Issues
Rakesh	1. Sending streaming data from phone to Spark. 2. Prediction of specific activity based on streaming data. 3. Labelling the clustered location data.	Develop, design, build and testing the task	100	22	
Nihar	1. Clustering location data using k-means. 2. Downtown location data collection.	Develop, design, build and testing the task	100	20	
Mark	1. Clustering location data using k-means. 2. Collecting location data set for clustering.	Develop, design, build and testing the task	100	22	
Sricharan	1. Socket connection between phone and spark. 2. Labelling the clustered data.	Develop, design, build and testing the task	100	20	

b. Work yet to be completed

Member	Task Description	Member Responsibility	Time (hours)	Comments /Issues
Rakesh	1. Implementing Machine Learning Algorithms in Spark R 2. Sending notification of predicted data. 3. Predicting the user patterns regularly. 4. Implementing API's.	Develop, design, build and testing the task	40	
Nihar	1. Implementing Machine Learning Algorithms in Spark R. 2. Collecting more sensor data sets. 3. Predicting the user patterns regularly. 4. Implementing API's.	Develop, design, build and testing the task	40	
Mark	1. Implementing Machine Learning Algorithms in Spark R 2. Sending notification of extracted data 3. Improving GUI. 4. Predicting the user patterns regularly.	Develop, design, build and testing the task	40	
Sricharan	1. Implementing Machine Learning Algorithms in Spark R 2. Sending notification of extracted data. 3. Testing the implemented features. 4. Predicting the user patterns regularly.	Develop, design, build and testing the task	40	

6. Bibliography:

i. Komninos, Andreas, and Mark Dunlop. "Text input on a smart watch." Pervasive Computing, IEEE 13.4 (2014): 50-58.

ii. Lutze, Rainer, and Klemens Waldhor. "A Smartwatch Software Architecture for Health Hazard Handling for Elderly People." Healthcare Informatics (ICHI), 2015 International Conference on. IEEE, 2015.

iii. http://www.motorola.com/us/products/moto-360

iv. https://www.thegrommet.com/robome-app-enabled-robot

v. https://www.yelp.com/developers/documentation/v2/overview

vi. https://api.eventful.com/