

Assist Robot

Project Report

Project Team – 8

Team Members

Priyadarsini Nidadavolu(17)

Deepthi Priyadarshini Penmetsa(22)

Dheeraja Vallabhaneni(28)

Tej Kumar Yentrapragada(33)

Table of Contents

Section	Description
1	Introduction
2	Project Goal and Objectives
3	Tools and Frameworks
4	Project Plan
	4.1 Schedule
	4.2 Project Timelines
	4.2.1 Team Members
	4.2.2 Tasks and Responsibilities
	4.3 Burndown Chart
5	Increment Report
	5.1 Increment based on Phases
	5.1.1 Phase 1 - Existing API
	5.1.2 Phase 2 - Recommendation System
	5.1.3 Phase 3 – Image Classification System
	5.2 Design of Features
	5.3 Implementation
	5.4 Deployment
6	Project Management
	6.1 Implementation Status Reports
7	Future Scope
8	Bibliography

1. Introduction:

The main goal of the project is to help people in finding their misplaced objects. Basically humans have a tendency to forget their belongings somewhere in their house and search for it for hours together. For example, if I have an important business meeting to attend, but I don't remember where I placed my car keys, then I might miss my meeting or can be late to the schedule. So to prevail in these circumstances here comes our Friendly Robot- My Friend which could assist me in keeping track of my personal things. So what this robot will do is that it will have entire map (laser scan) of the building and objects in the building in its memory. So we will feed the robot with the objects that are highly important to us, like car keys, some files, phone and laptop. This robot will keep its eye on these objects and notify their location to its master upon request. Additionally our Robot who will be an eFriend who will help us to choose the furniture to our home. Also our robot will suggest us the top rated books.

2. Project Goal and Objectives:

The primary goals of our project is described below:

- To implement a module to create an android application where the user is provided with the login page. The user will be provided with the features of registering the new account or to login with an existing account.
- To implement an android module where user identification is done using facial reorganization.
- To implement an android module where there must be a camera facility. The objects image must be captured and should save the image to the SD card.
- To implement a module which could implement client server module where the client (android device) will be sending the streaming images to server (Spark).
- To implement an image classification module using machine learning techniques in which the image that comes from the android device should be classified.
- To implement a recommendation module where the location of the classified image should be recommended.
- To implement a module where the notifications about the classification and recommendation will be sent to an android device.

- To build a recommendation system which will suggest us about the latest furniture details, their quality and from which brand/ shop we could purchase from. This feature enables us to decorate our houses with rich interior designing.

3. Tools and Frameworks

Android Studio:

An official Integrated Developer Environment which we used for Android application development. It is based on IntelliJ IDEA with powerful code editor and developer tools embedding flexible ‘Gradle’ based build system. It also support drag and drop editing which helped us in developing rich User Interface. It also had built in feature like emulator where we can test our developed applications without being deployed on actual device.

IntelliJ IDEA:

It's a Java Integrated Development Environment which helps to develop software applications with the support of various languages like Scala, SQL, Java which we had used for our project. It provides variety of build and packaging tools like Maven, Grunt, Gradle, SBT etc.

Maven – A build automation tool developed by Apache organization that is used for adding dependencies and external JARS for JAVA based projects.

SBT – An open source tool which can run on cross platform for building Scala and JAVA based projects. An added advantage of this build tool is that it can compile Scala projects by integrating them with many test frameworks. It has a special feature called Ivy with which it provides dependency management.

Spark:

Spark is an open source framework that runs on top of Hadoop clustering systems. It is experimentally proved that Spark can process the data at least 10 times faster than when it runs on local disk. Spark has a variety of features and it supports various languages like R, SQL, Python, Scala and JAVA. In the project we had used Spark features like MLlib, streaming and data frames. Spark MLlib is a Machine Learning library which has the algorithms like Naive Bayes, Random Forest, and Alternate Least Squares etc.

4. Project Plan:

4.1 Schedule:

Stories: Four user stories had been created as part of Iteration 1. Here are the snapshots for the stories which are in closed and opened state.

This screenshot shows the GitHub Issues page for the repository 'npdarsini/Assist-Robot'. The search bar contains the query 'is:open issue author:npdarsini'. The results show two open issues:

- Test - Different Objects [enhancement] #4 opened 35 minutes ago by npdarsini ↑ Project - Increment 1
- Capture and feed the Images [enhancement] #3 opened 38 minutes ago by npdarsini ↑ Project - Increment 1



This screenshot shows the GitHub Issues page for the repository 'npdarsini/Assist-Robot'. The search bar contains the query 'is:closed issue author:npdarsini'. The results show two closed issues:

- Smart Watch connection with Smart Phone [enhancement] #2 opened 40 minutes ago by npdarsini ↑ Project - Increment 1
- Smart Phone connection with Android Studio [enhancement] #1 opened 42 minutes ago by npdarsini ↑ Project - Increment 1



Stories: Eight user issues had been created as part of Iteration 2. Here are the snapshots for the stories which are in closed and opened state.

Issues | Pull requests | Issues | Gist | ToDo

npdarsini / Assist-Robot

Issues

Filters: Is issue milestone: "Project - Increment 2" is closed

1 Open ✓ 7 Closed

- Creating a perfect training Dataset and collecting Test data (enhancement)
- Usage of Speech and image services (enhancement)
- Data Collection and analysis (enhancement)
- Establishing connection between smart watch and Robome (enhancement)
- Installation of Spark (enhancement)
- Features of Robome (question)
- Introduction to Spark (enhancement)

ProTip! Adding no label will show everything without a label.



Issues | Pull requests | Issues | Gist | ToDo

npdarsini / Assist-Robot

Issues

Filters: Is issue milestone: "Project - Increment 2" is open

1 Open ✓ 7 Closed

- Developing a Recommendation System (enhancement)

ProTip! Mix and match filters to narrow down what you're looking for.



Stories: Six user issues had been created as part of Iteration 3. Here are the snapshots for the stories which are in closed and opened state.

The screenshot shows the GitHub Issues page for the repository npdarsini/Assist-Robot. There are 17 issues listed:

- Notification to Android Device - Recognized Object (enhancement) #18 (closed)
- Training data sets - Creation with different objects (enhancement) #17 (closed)
- Exploring different Classification algorithms - Random Forest, Decision Tree (enhancement) #16 (closed)
- Image classification (enhancement) #15 (closed)
- Developing a Recommendation System (enhancement) #14 (closed)
- Introduction to ML Algorithm (enhancement) #13 (closed)
- Creating a perfect training Dataset and collecting Test data (enhancement) #11 (closed)
- Usage of Speech and image services (enhancement) #10 (closed)
- Data Collection and analysis (enhancement) #9 (closed)
- Establishing connection between smart watch and Robome (enhancement) #8 (closed)
- Linking Modules (enhancement) #20 (open)
- Classification based on the Streaming Data (enhancement) #19 (open)
- Streaming Data Collection (enhancement) #12 (open)

The screenshot shows the GitHub Issues page for the repository npdarsini/Assist-Robot. There are 3 open issues:

- Linking Modules (enhancement) #20
- Classification based on the Streaming Data (enhancement) #19
- Streaming Data Collection (enhancement) #12

The screenshot shows a Windows taskbar with three open files:

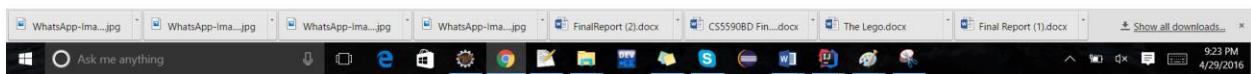
- FinalReport.docx
- Train.zip
- modelZip

Stories:

This screenshot shows the GitHub Issues page for the repository npdarsini/Assist-Robot. The search bar at the top contains the query "is:issue is:open". The results list six open issues:

- #23: classifying images with images captured through android device
- #22: receiving streamed data to spark
- #21: Streaming data from android
- #20: Linking Modules enhancement
- #19: Classification based on the Streaming Data enhancement
- #12: Streaming Data Collection enhancement

The interface includes standard GitHub navigation elements like filters, labels, milestones, and a "New issue" button.

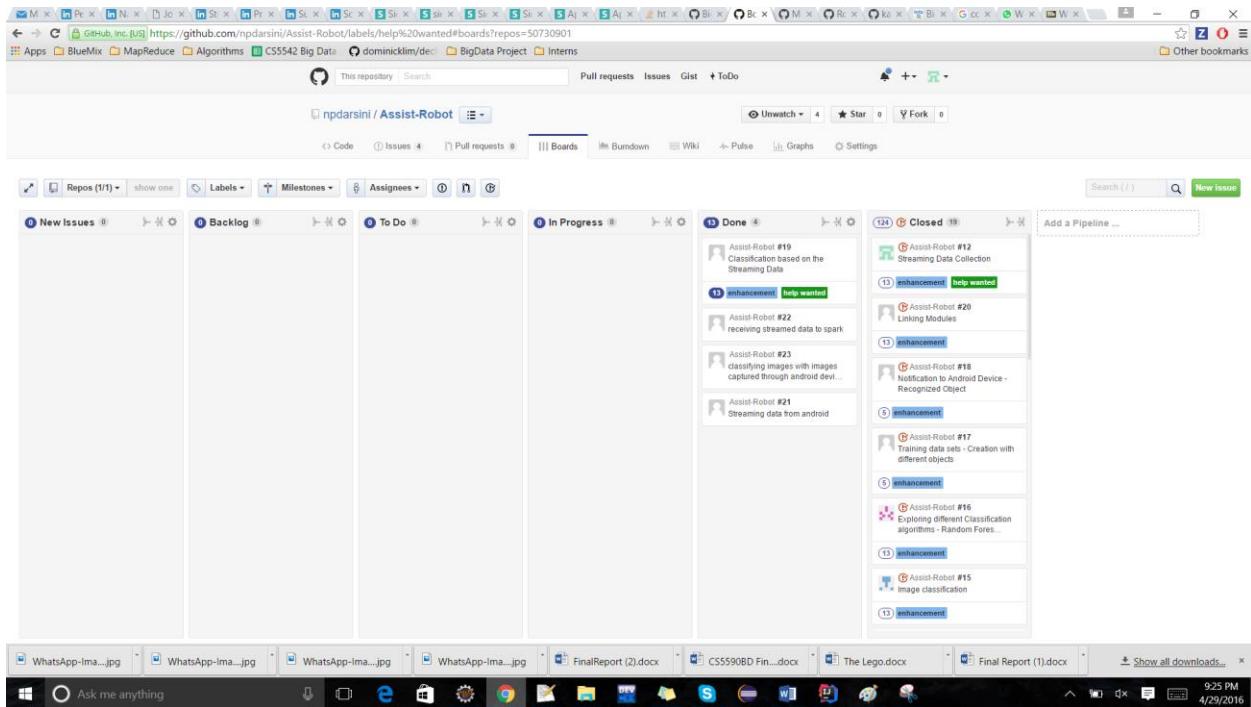


This screenshot shows the GitHub Issues page for the repository npdarsini/Assist-Robot. The search bar at the top contains the query "is:issue is:closed". The results list 23 closed issues:

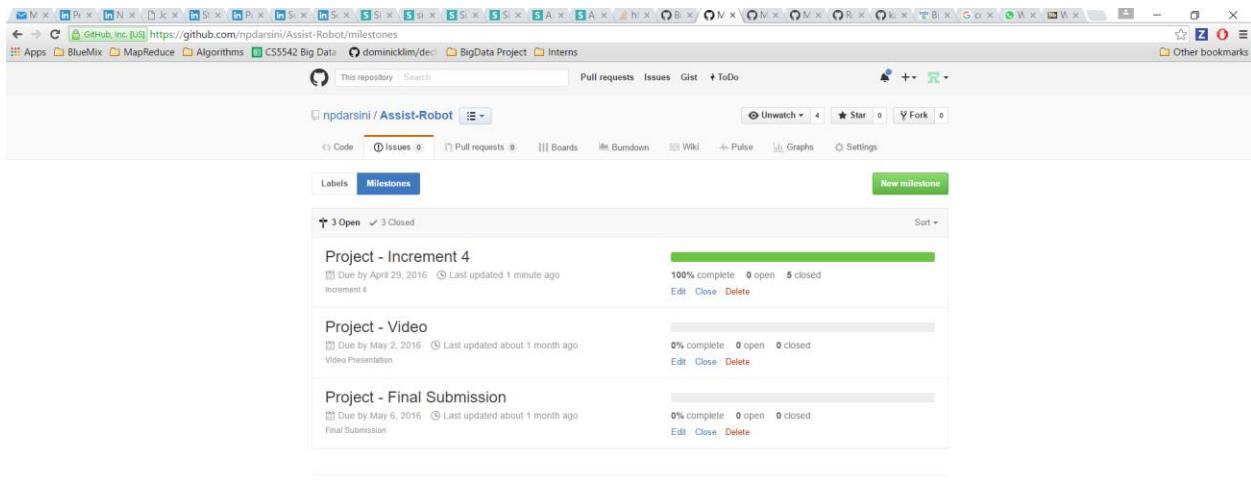
- #23: classifying images with images captured through android device
- #22: receiving streamed data to spark
- #21: Streaming data from android
- #20: Linking Modules enhancement
- #19: Classification based on the Streaming Data enhancement
- #18: Notification to Android Device - Recognized Object enhancement
- #17: Training data sets - Creation with different objects enhancement
- #16: Exploring different Classification algorithms - Random Forest, Decision Tree enhancement
- #15: Image classification enhancement
- #14: Developing a Recommendation System enhancement

The interface includes standard GitHub navigation elements like filters, labels, milestones, and a "New issue" button.

Board:

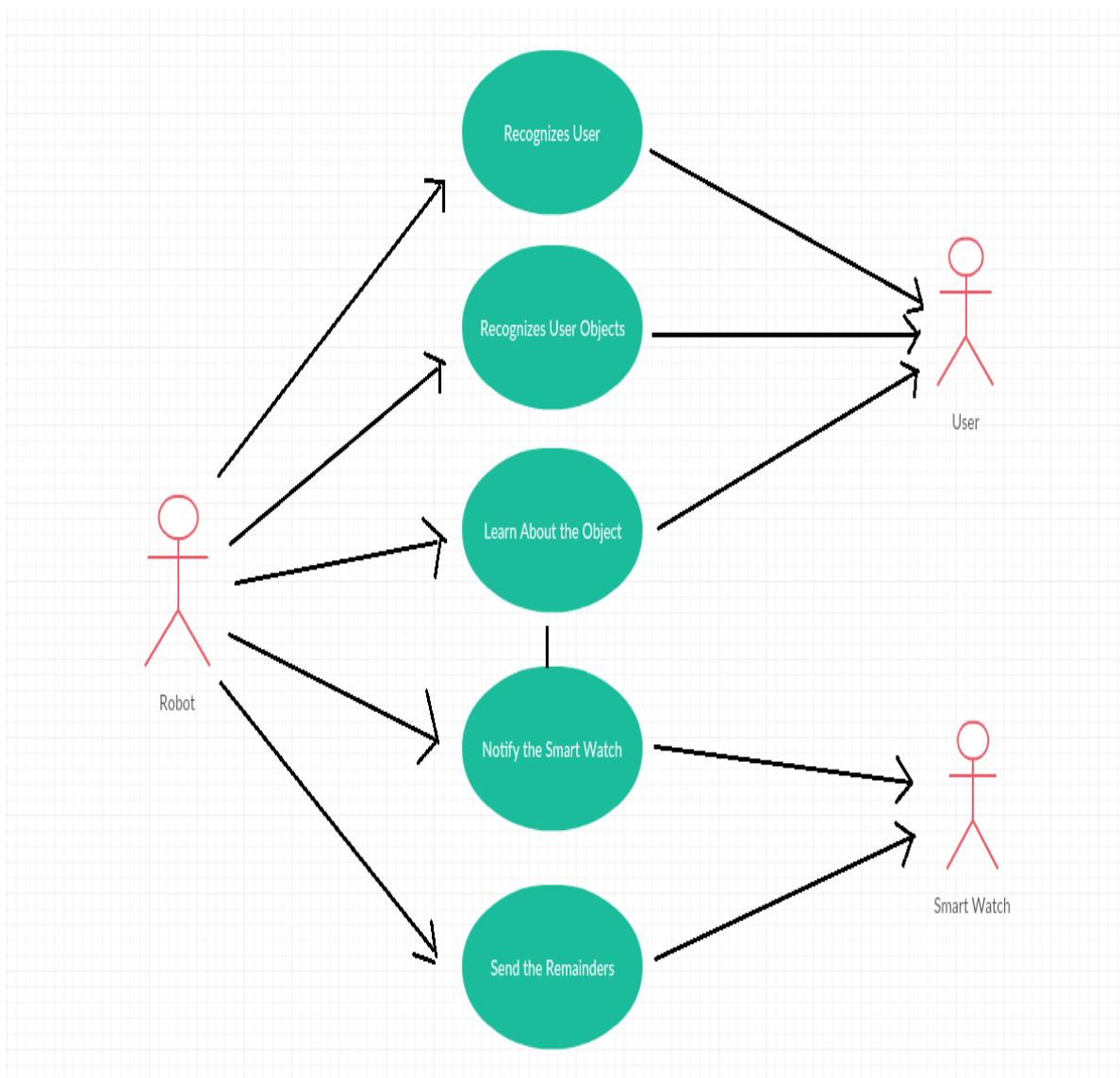


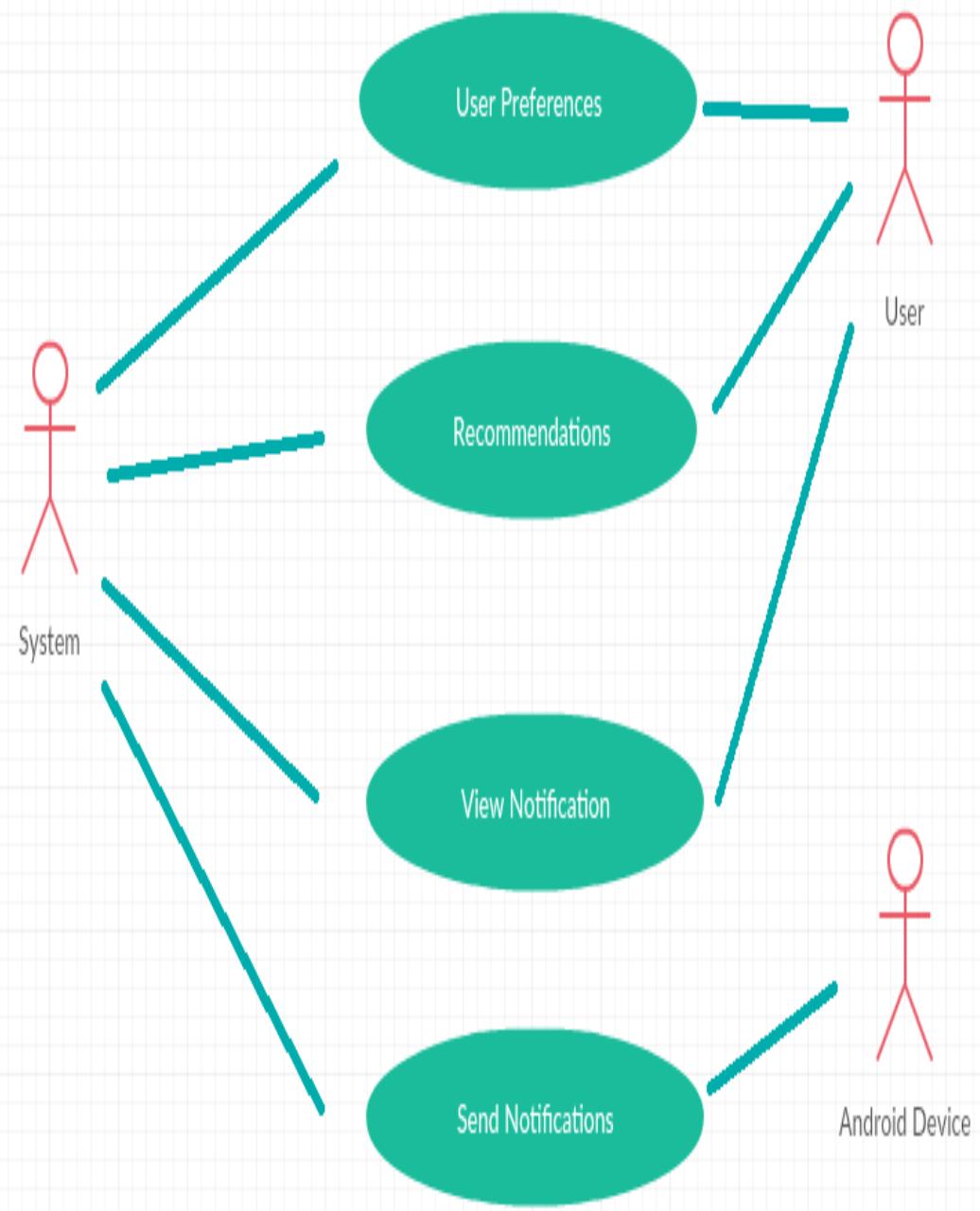
Milestones:



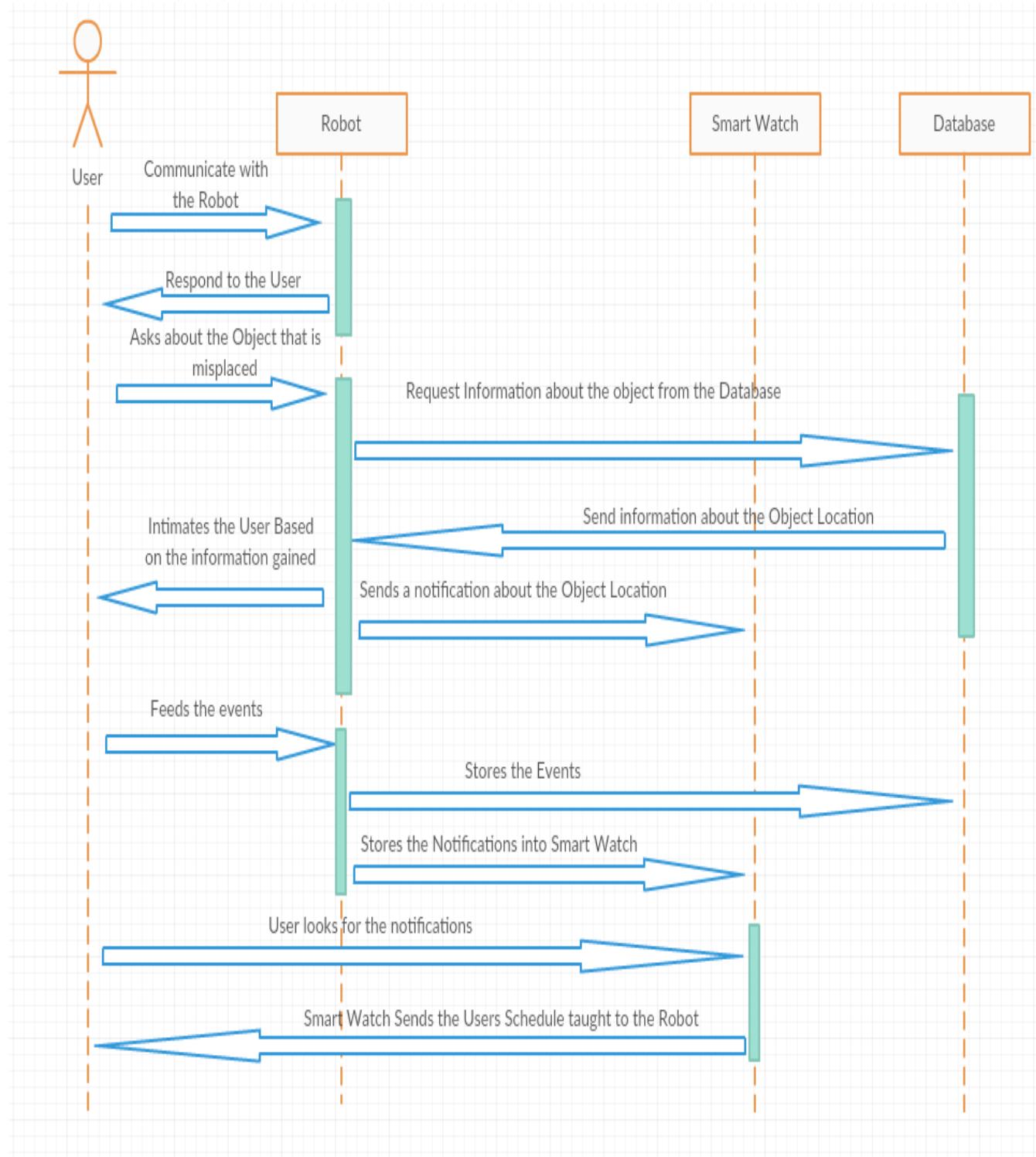
4.1.1 UML Diagrams:

Use Case Diagram

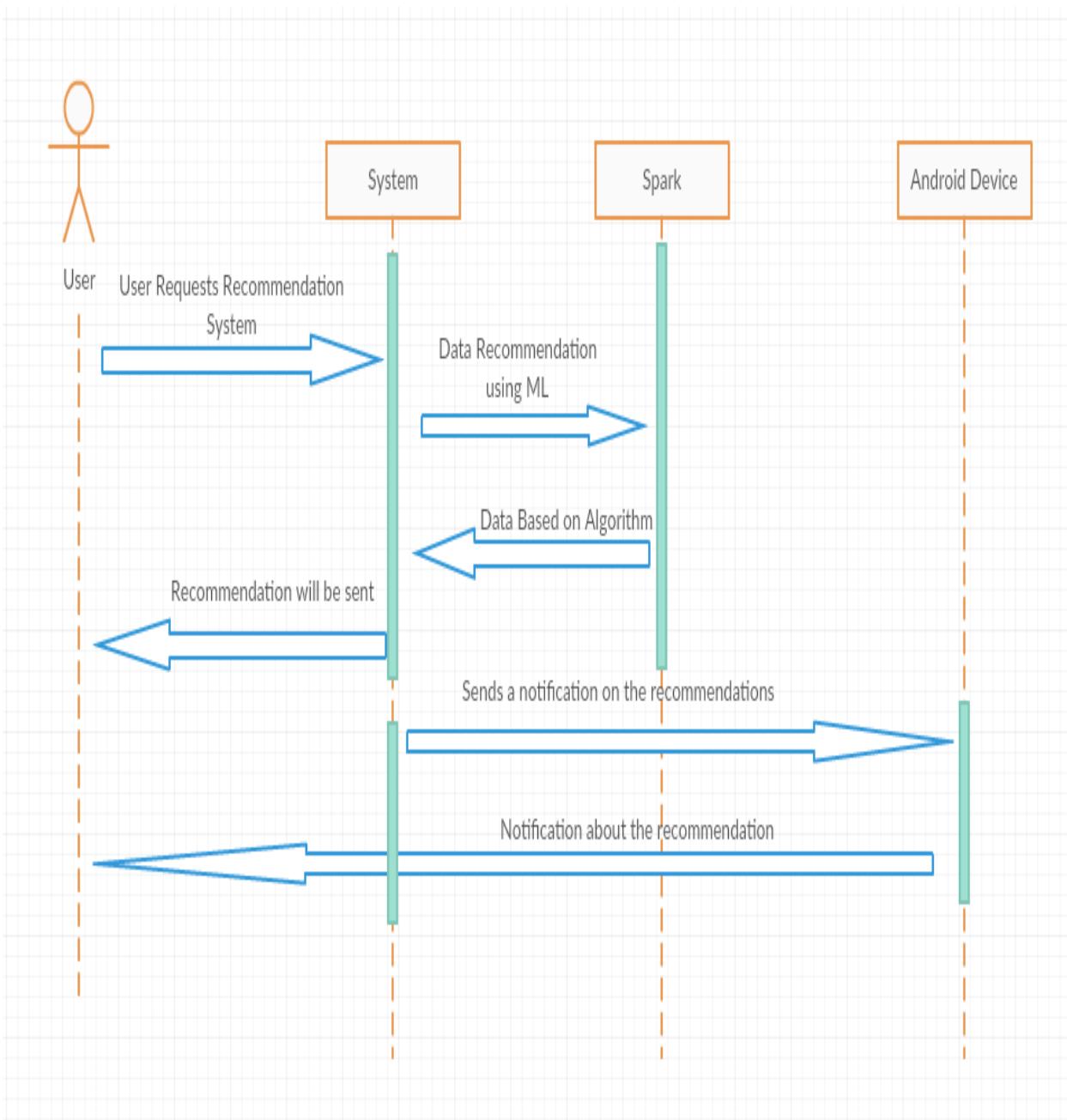




Sequence Diagram:



Sequence Diagram for Recommendation System



4.2 Project Timelines:

Increment	Deadline
Increment 1	19 February 2015
Increment 2	11 March 2016
Increment 3	6 April 2016
Increment 4	29 April 2016
Final Submission	6 May 2016

4.2.1 Team Members:

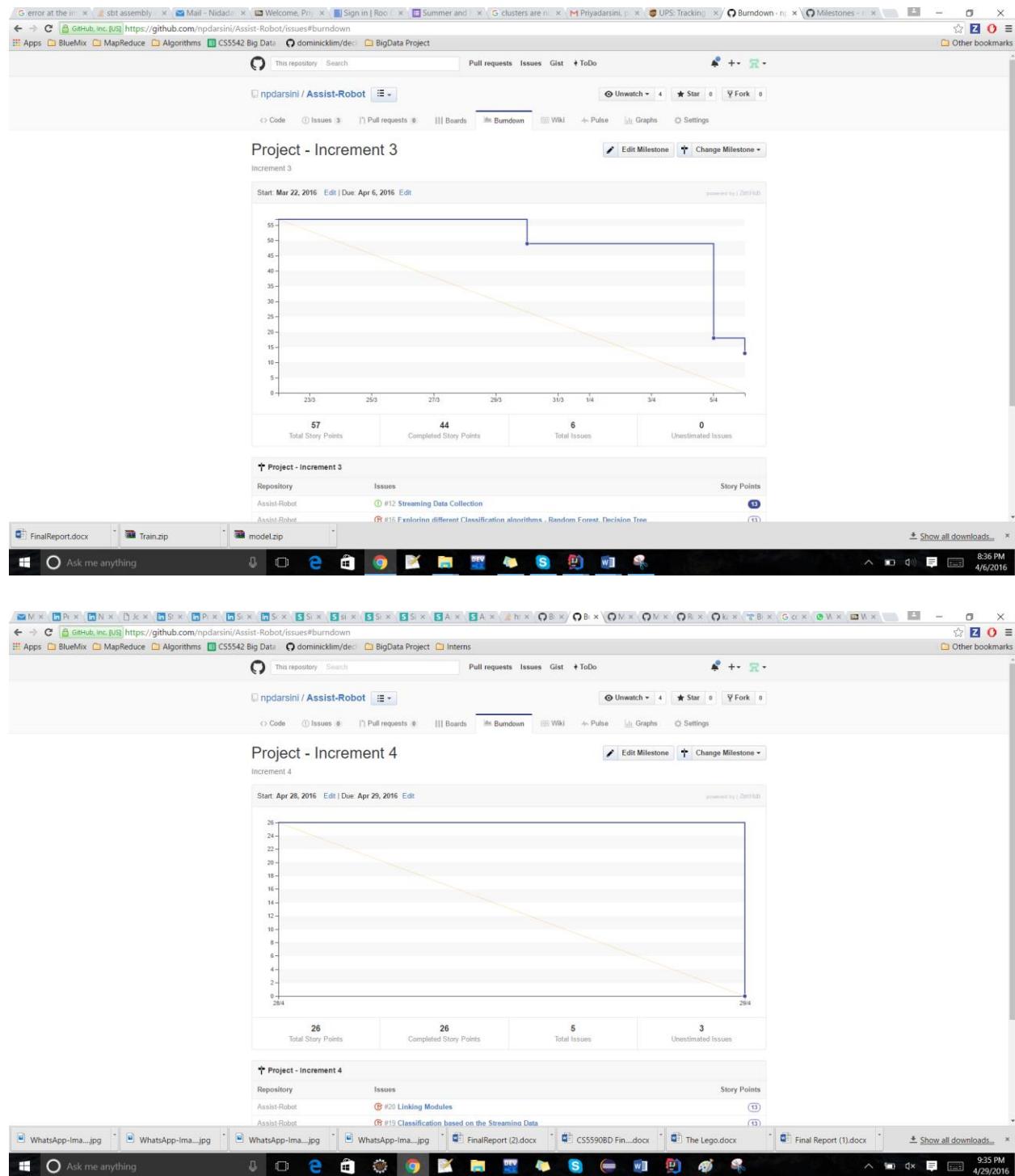
1. Priyadarsini Nidadavolu – 17
2. Deepthi Priyadarshini Penmetsa – 22
3. Dheeraja Vallabhaneni – 28
4. Tej Kumar Yentrapragada – 33

4.2.2 Tasks and Responsibilities:

Name	Responsibilities
Priyadarsini Nidadavolu	Machine Learning with Spark
Deepthi Priyadarshini Penmetsa	Spark and Hadoop Technologies
Tej Kumar Yentrapragada	Android Programming and Spark MLLib
Dheeraja Vallabhaneni	Android Programming

4.3 Burndown Chart:

Burndown:



5. Increment Report

5.1 Incremental Explanations

5.1.1 Phase 1 -Existing API:

IBM Alchemy API

This API basically performs machine learning and natural language processing techniques. Some of its features include semantic text analysis, sentimental analysis, deep learning, face detection and reorganization, speech to text and vice versa conversions etc. In this we had used this API in order to recognize the objects that we want to teach the Robot.

Achievements upon using this API – The Robot could identify basic objects like laptop, phone, bottle etc.

5.1.2 Phase 2 - Recommendation System:

In this phase we had developed two recommendation systems which can recommend the user about the popular furniture showrooms and the famous books. In this we provided the training set with user information (uid, name, ratings etc.,), furniture information (list of showrooms, location) and the book information (name, author etc.,).

The recommended notification has been sent to the android device (smart watch/phone) using Spark-Android Socket programming techniques.

5.1.3 Phase 3 – Image Classification:

In this phase we had implemented Image Classification system using Random Forest Machine Learning Classifier Algorithm. In this we had provided the training data set which has the different kinds of objects like keys, charger, watch, spectacles, phone. Basically, we generated the key descriptors and created the clusters and histograms out of it. Based on this features, the classifier predicts the image from the testing data set that has been provided.

The notification about the classified image will be sent to the android device (smart watch/phone) using Spark-Android Socket programming techniques.

5.1.4 Final Phase

This phase basically projects full pledged implementation of our project.

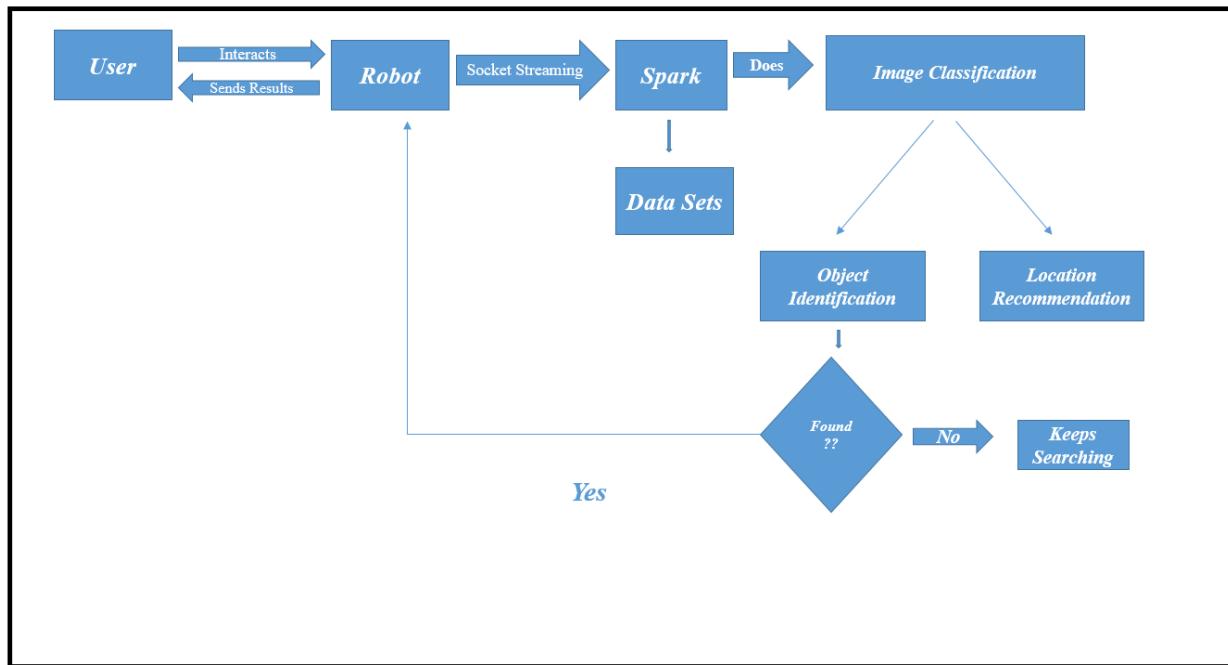
Login to the android application either by using Email ID as a user id or by facial recognition. The user is provided to register his own account or can be able to login with the existing signon. The login page is developed using Java in Android Studio IDE. Upon logging in, the camera option will be made available to the user. Android device will be capturing the images and store it in the SD card using android.os.Environment library.

A client server module has been developed to send these streaming images to spark testing directory. The Spark Server will accept the connection from android client and will receive the base64 formatted image string. At the Spark Server side, these images will be decoded and classified using Random Forest algorithm. This classified result will be send to the user device in the form of notification. Based on the image that is classified, a recommendation system has been built. If the object is out of reach, it will give the recommended places that the object can be placed based on the training data. This will also be sent to the user in the form of notification.

5.2 Design of Features:

The architecture of our system could be like the user can login to the android application with the facial recognition or with the user credentials. Image capture could be done and will be sent to spark as a streaming data via socket. Image Classification based on the training and testing data will be done and further a recommendation system has been developed. The outcomes from these algorithms will be sent to the android devices as a notification.

Work Flow:



System Features

The following are the features that were developed as part of Phase I:

We had used IBM's Alchemy API and able to make our Robot to detect the object and return the object name as a result.

The following are the features that are developed as part of Phase II:

We had used machine learning algorithms to develop a recommendation system. In this phase we had developed two recommendation systems with which the system will be able to suggest top rated books to the user based on his interests and the furniture showrooms which could be available for cheaper prices with the location. Basically, we had provided the training data sets and the user preferences which serves as a key inputs for the system. We were also able to connect our system to the android device to which the recommendations has been sent.

The following are the features that are developed as part of Phase III:

We had used Random Forest Classifier Algorithm to develop an Image classification system. In this phase we had provided the system with the training data set which consists the sample images of the different types of objects like keys, watch, spectacles, phone etc. We are able to create the clusters and histograms out of the provided data and was able to send a notification to the smart device about the object that has been predicted.

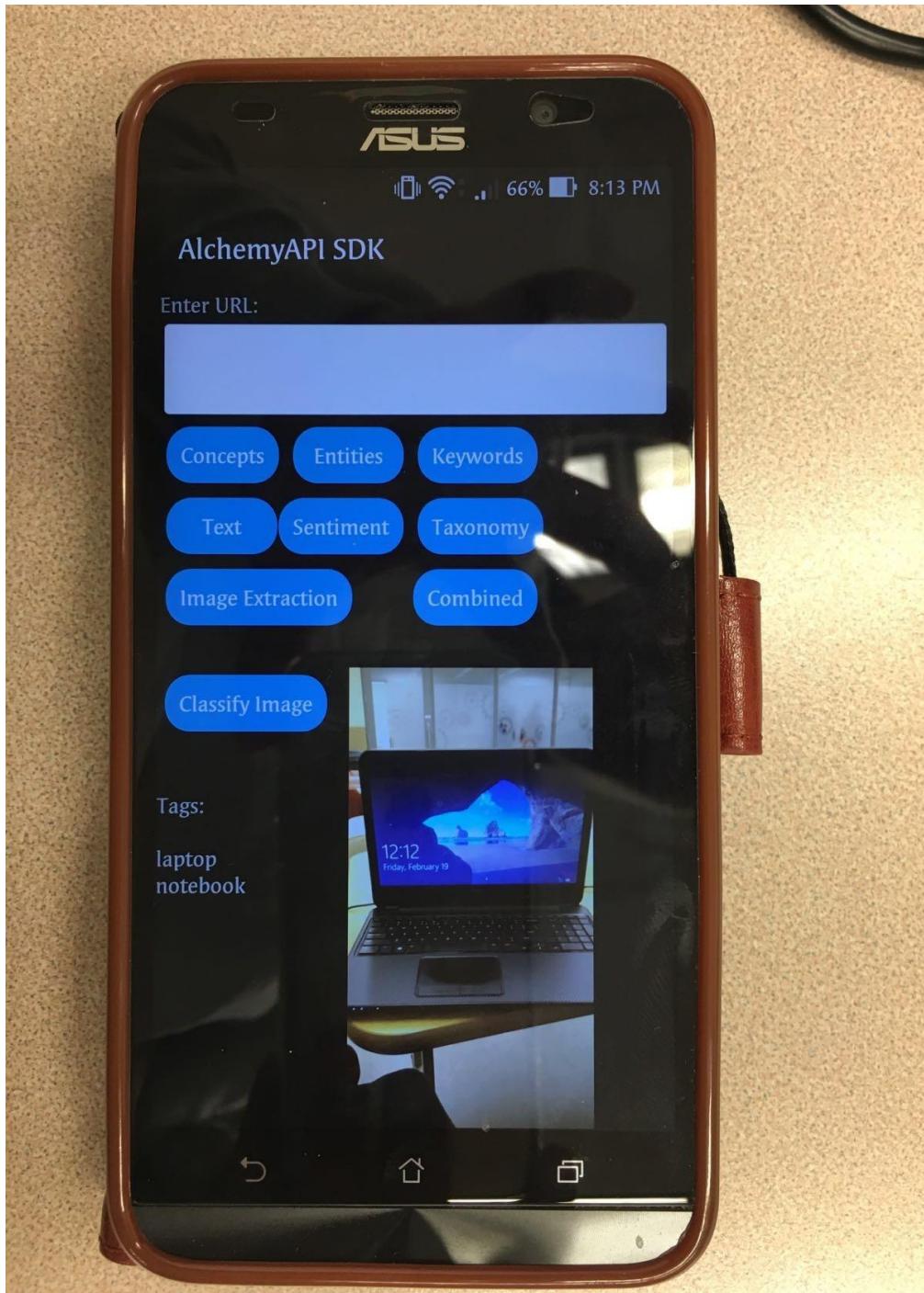
The following features has been added for the final phase:

In the previous phase Image classification is done with the static testing data set. In this phase we had added the functionality of streaming image classification. In this the streaming data will be coming from the android device to the Spark server. A login page for the android application has been developed and embedded with the functionality of facial recognition and user registration options. Upon logging in to the application the image capture option is implemented with the facility to store it to the device SD card.

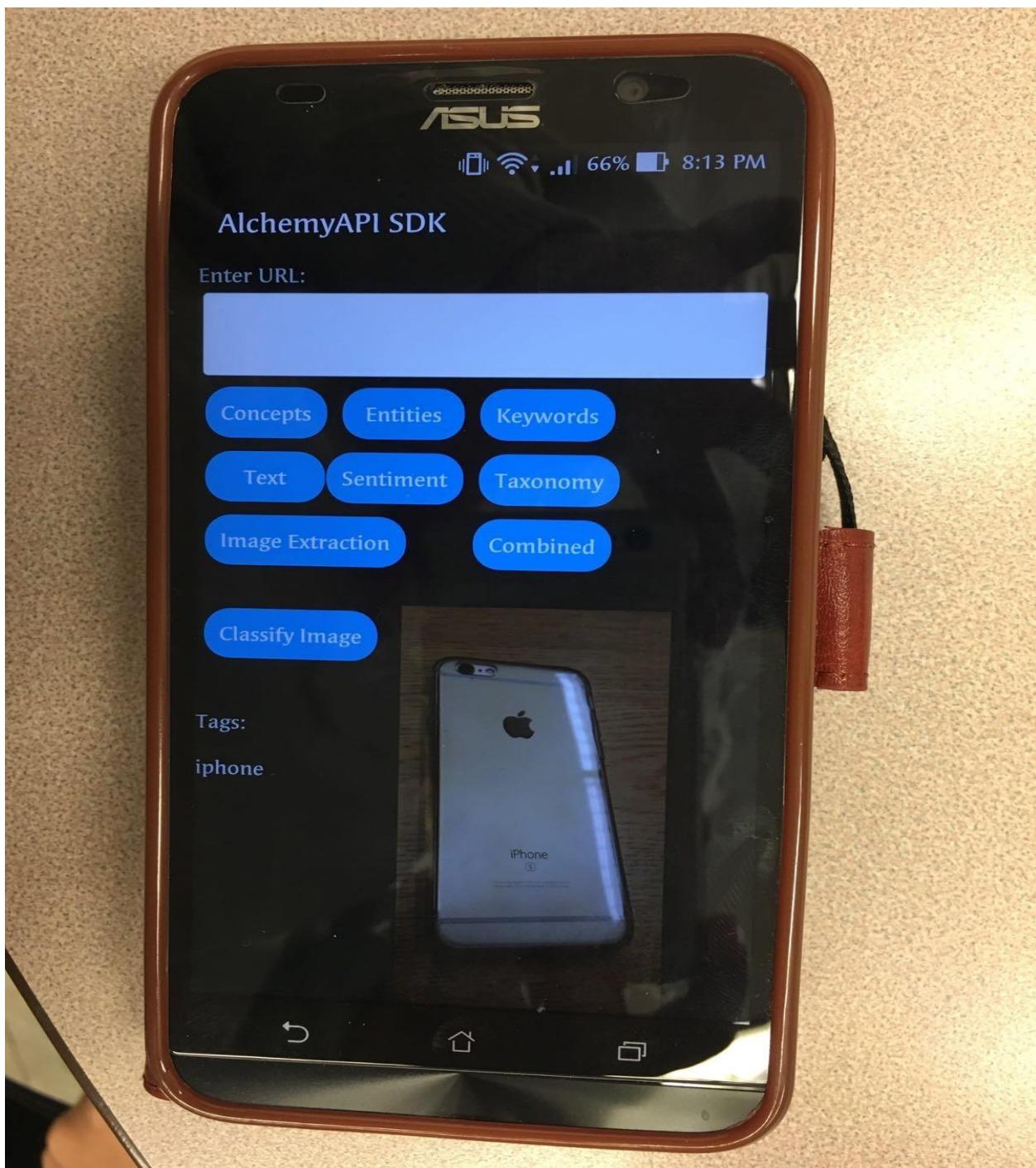
5.3 Implementation:

Mobile Client Implementation – Snapshots

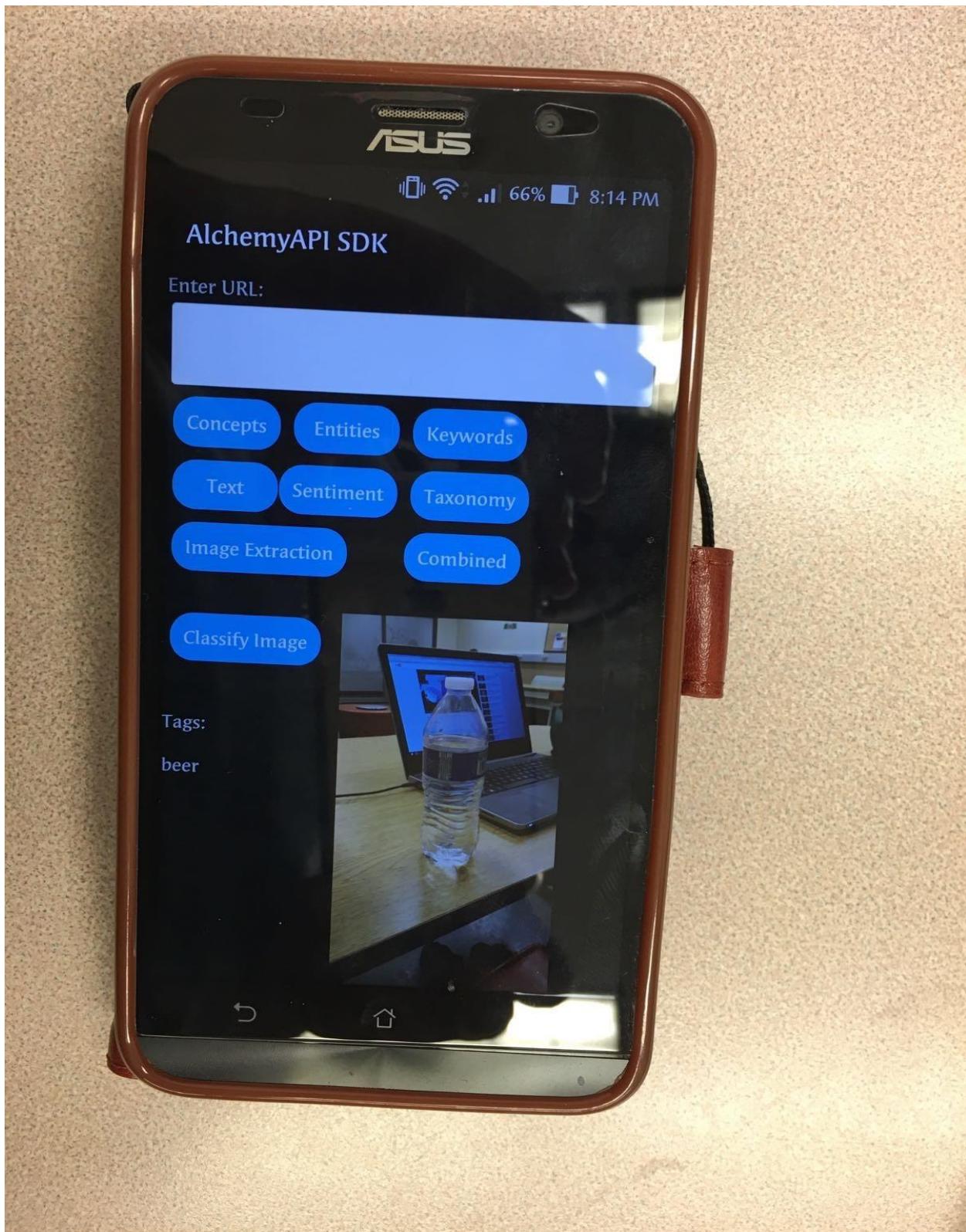
This snapshot shows us that the application is able to identify the object and names its Laptop.



This snapshot shows us that the application is capable of identifying the object and names it as an Iphone.



This snapshot shows us that the application is able to identify the bottle.



Recommendation System Snapshots:

Books Recommendation System:

Phase at which the analyzation of training data set is taking place.

SparkRecommendationSystem - C:\Users\npdar\Desktop\Acads\BigDataAnalytics\Assignment\Tutorial 7\RecommendationSystem\SparkMachineLearnin - [sparkmachinelearnin] - ...src\main\scala\com\umkc\sparkML\SocketClient.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkMachineLearnin [sparkmachinelearnin] (C:\Users\npdar\...)

Project Books Books.dat

SparkCollaborativeFiltering.scala personalRating.txt SparkBookRecommendation.scala users.dat Books.dat SocketClient.scala

Run SparkBookRecommendation

lazy val address: Array[Byte] = Array(10.toByte, 99.toByte, 0.toByte, 203.toByte)
val ia = InetSocketAddress.getByAddress(address)
val socket = new Socket(ia, 1234)

at org.apache.spark.SparkContext.runJob(SparkContext.scala:1929)
at org.apache.spark.rdd.RDD.\$anonfun\$collect\$1(RDD.scala:1923)
at org.apache.spark.rdd.RDDOperationScope\$.withScope(RDDOperationScope.scala:159)
at org.apache.spark.rdd.RDD.\$anonfun\$collect\$1(RDD.scala:1923)
at org.apache.spark.rdd.RDD.withScope(RDD.scala:111)
at org.apache.spark.rdd.RDD.collect(RDD.scala:1530)
at com.umkc.sparkML.SparkBookRecommendation\$.main(SparkBookRecommendation.scala:56)
at com.umkc.sparkML.SparkBookRecommendation.main(SparkBookRecommendation.scala) <5 internal calls>

16/03/11 20:23:39 WARN MetricsSystem: Your hostname, LAPTOP-FK785LNH resolves to a loopback/non-reachable address: fe80:0:0:f9c9:4c21:d406:39d2%eth4, but we couldn't find any external IP.

16/03/11 20:23:40 INFO FileInputFormat: Total input paths to process : 1

16/03/11 20:23:41 INFO Deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id

16/03/11 20:23:41 INFO Deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id

16/03/11 20:23:41 INFO Deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.id

16/03/11 20:23:41 INFO Deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

16/03/11 20:23:41 INFO Deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

16/03/11 20:23:41 INFO Deprecation: mapred.task.id.map is deprecated. Instead, use mapreduce.task.iemap

16/03/11 20:23:41 INFO Deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition

16/03/11 20:23:41 INFO Deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id

16/03/11 20:23:41 INFO FileInputFormat: Total input paths to process : 1

Got 100020 ratings from 604 users on 3706 Books.

Training: 602252, validation: 189919, test: 199049

[Stage 25]: (0 + 4) / 4116/03/11 20:23:59 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.Native

16/03/11 20:24:01 INFO JniLoader: successfully loaded C:\Users\npdar\AppData\Local\Temp\jniloader20562222124190243netlib-native_ref-win-x86_64.dll

16/03/11 20:24:02 WARN LAPACK: Failed to load implementation from: com.github.fommil.netlib.NativeSystemLAPACK

16/03/11 20:24:02 INFO JniLoader: already loaded netlib-native_ref-win-x86_64.dll

[Stage 163]: (0 + 4) / 4)RMSE (validation) = 0.881176457356829 for the model trained with rank = 8, lambda = 0.1, and numIter = 20.

RMSE (validation) = 0.8728702475180768 for the model trained with rank = 8, lambda = 0.1, and numIter = 20.

RMSE (validation) = 3.7558695311242833 for the model trained with rank = 8, lambda = 10.0, and numIter = 10.

RMSE (validation) = 3.7558695311242833 for the model trained with rank = 8, lambda = 10.0, and numIter = 20.

RMSE (validation) = 0.8778839053170892 for the model trained with rank = 12, lambda = 0.1, and numIter = 10.

RMSE (validation) = 0.8714205258125456 for the model trained with rank = 12, lambda = 0.1, and numIter = 20.

[Stage 1256]: (3 + 1) / 4]

Event Log 16

Find Run Debug TODO Messages Terminal

All files are up-to-date (2 minutes ago)

7081 CRLF UTF-8

825 3/11/2016

Recommended Books

SparkRecommendationSystem - [C:\Users\ndpar\Desktop\Acads\BigDataAnalytics\Assignment\Tutorial 7\Recommendation System\SparkMachineLearnin - [sparkmachinelearnin - \src\main\scala\com\umkc\sparkML\SocketClient.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkMachineLearnin > src > main > scala > com > umkc > sparkML > SocketClient.scala

SparkCollaborativeFiltering.scala x personalRating.txt x SparkBookRecommendation.scala x users.dat x Books.dat x SocketClient.scala x SparkBookRecommendation -

Project Structure

SparkMachineLearnin [sparkmachinelearnin] C:\Users\ndpar\...
Books.dat

Run SparkBookRecommendation

[Stage 1634:=====]
[Stage 1685:
The best model improves the baseline by 21.91%.
Books that are recommended for you:
1: Flood: Mississippi 1927
2: The Middle Stories
3: Jane Doe
4: Wild Animals
5: Where You'll Find Me: And Other Stories
6: Tell Me This Isn't Happening
7: Goodbye to the Buttermilk Sky
8: Hitler's Secret Bankers: The Myth of Swiss Neutrality During the Holocaust
9: What If?: The World's Foremost Military Historians Imagine What Might Have Been
10: PLEADING GUILTY
11: Clara Callan
12: The Mummies of Urucumhi
13: The Testament
14: Nights Below Station Street
15: A Second Chicken Soup for the Woman's Soul (Chicken Soup for the Soul Series)
16: New Vegetarian: Bold and Beautiful Recipes for Every Occasion
17: Under the Black Flag: The Romance and the Reality of Life Among the Pirates
18: Timeline
19: More Cunning Than Man: A Social History of Rats and Man
20: Airframe
21: The Witchfinder (Amos Walker Mystery Series)
22: The Kitchen God's Wife
23: Our Dumb Century: The Onion Presents 100 Years of Headlines from America's Finest News Source
24: Beloved (Plume Contemporary Fiction)

Process finished with exit code 0

Find Run Debug TODO Messages Terminal Event Log

All files are up-to-date (11 minutes ago)

87:18 CRLF- UTF-8 3/11/2016 8:34 PM

We had also sent the recommended books to the smart phone as a notification:



Furniture Malls Recommendation System

```
16/03/10 07:16:28 INFO FileInputFormat: Total input paths to process : 1
Got 1000209 ratings from 6040 users on 3706 FurnitureMalls.
Training: 602252, validation: 198919, test: 199049
[Stage 25:>                                         (0 + 4) / 4]16/03/10 07:16:25 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeDenseMatrix64F
16/03/10 07:16:28 INFO JniLoader: successfully loaded C:\Users\DEEPU\AppData\Local\Temp\jniloader1966739238328846810netlib-native_ref-win-x86_64.dll
[Stage 27:>                                         (0 + 4) / 4]16/03/10 07:16:33 WARN LAPACK: Failed to load implementation from: com.github.fommil.netlib.NativeDenseMatrix64F
16/03/10 07:16:33 INFO JniLoader: already loaded netlib-native_ref-win-x86_64.dll
RMSE (validation) = 0.8815801121709103 for the model trained with rank = 8, lambda = 0.1, and numIter = 10.
RMSE (validation) = 0.8726203182715503 for the model trained with rank = 8, lambda = 0.1, and numIter = 20.
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 8, lambda = 10.0, and numIter = 10.
[Stage 950:>                                         (0 + 4) / 4]RMSE (validation) = 3.7558695311242833 for the model trained with rank = 8, lambda = 10.0, and numIter = 20.
[Stage 1125:=====                                         (3 + 1) / 4]RMSE (validation) = 0.8772284010651425 for the model trained with rank = 12, lambda = 0.1, and numIter = 20.
RMSE (validation) = 0.8710227453579589 for the model trained with rank = 12, lambda = 0.1, and numIter = 20.
[Stage 1595:>                                         (0 + 4) / 4]RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 20.
RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 20.
The best model was trained with rank = 12 and lambda = 0.1, and numIter = 20, and its RMSE on the test set is 0.8690494992690084.
The best model improves the baseline by 21.95%.
```

This Screenshot shows the Furniture Malls recommended to you:

The screenshot shows the IntelliJ IDEA 15.0.1 interface with the following details:

- Project Structure:** Shows the project structure with files like SparkCollaborativeFiltering.scala, SparkMovieRecommendation.scala, FurnitureShopRating.txt, sample_libsvm_data.txt, and test.data.txt.
- Code Editor:** Displays the Scala code for `SparkMovieRecommendation`. The code includes a call to `sys.exit(1)` at the end.
- Run Output:** The terminal pane shows the execution logs:
 - INFO deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition
 - INFO deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id
 - INFO FileInputFormat: Total input paths to process : 1
 - Got 1000209 ratings from 6040 users on 3706 FurnitureMalls.
 - Training: 602252, validation: 198919, test: 199049
 - [Stage 25:> (0 + 4) / 4]16/03/10 07:16:25 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeDenseMatrix64F
 - [Stage 27:> (0 + 4) / 4]16/03/10 07:16:33 INFO JniLoader: successfully loaded C:\Users\DEEPU\AppData\Local\Temp\jniloader1966739238328846810netlib-native_ref-win-x86_64.dll
 - [Stage 950:> (0 + 4) / 4]RMSE (validation) = 3.7558695311242833 for the model trained with rank = 8, lambda = 10.0, and numIter = 20.
 - [Stage 1125:===== (3 + 1) / 4]RMSE (validation) = 0.8772284010651425 for the model trained with rank = 12, lambda = 0.1, and numIter = 20.
 - [Stage 1595:> (0 + 4) / 4]RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 20.
 - RMSE (validation) = 3.7558695311242833 for the model trained with rank = 12, lambda = 10.0, and numIter = 20.
 - The best model was trained with rank = 12 and lambda = 0.1, and numIter = 20, and its RMSE on the test set is 0.8690494992690084.
 - The best model improves the baseline by 21.95%.
- Bottom Status Bar:** Shows "Compilation completed successfully in 31s 964ms (7 minutes ago)" and the current time as 43:1 CRLF: UTF-8:.

The notification has been sent to the android mobile which shows the recommended furniture list.

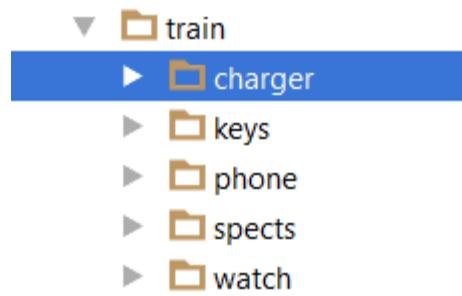


Image Classification

The Training dataset is full with images of objects the user want to identify which were got misplaced. And the test dataset is the streaming images which are captured by the robot while trying to identify the misplaced object. So when the robot finds the image, it classifies the image and notifies about the object to the users device.

Screenshots of the image classification:

Train data:



Identifying Key descriptors from Training dataset

```
// Empty categoricalFeaturesInfo indicates all features are continuous.
val numClasses = 10
val categoricalFeaturesInfo = Map[Int, Int]()
// val numTress = 10 // Use more in practice.
// val featureSubsetStrategy = "auto" // Let the algorithm choose.
// val impurity = "gini"
// val maxDepth = 4
val maxBins = 100
```

Key Descriptors 1556 x 128
Key Descriptors 1334 x 128
Key Descriptors 1000 x 128
Key Descriptors 692 x 128
Key Descriptors 515 x 128
Key Descriptors 781 x 128
Key Descriptors 20587 x 128
Key Descriptors 24357 x 128
Key Descriptors 2381 x 128
Key Descriptors 24357
Key Descriptors 17111 x 128
Key Descriptors 27961 x 128
Key Descriptors 3518 x 128
Key Descriptors 27961

Cluster formation:

The screenshot shows the IntelliJ IDEA interface with the 'Image_Classification' project open. The 'IPApp.scala' file is the active editor, containing Scala code for a machine learning model. The 'Run' tool window displays the application's log output, which includes various INFO and WARN messages related to the file input format and task context. The status bar at the bottom right indicates the date as 4/6/2016.

```
// Empty categoricalFeaturesInfo indicates all features are continuous.
val numClasses = 10
val categoricalFeaturesInfo = Map[Int, Int]()
// val numTrees = 10 // Use more in practice.
// val featureSubsetStrategy = "auto" // Let the algorithm choose.
// val impurity = "gini"
// val maxDepth = 4
val maxBins = 100
```

Logs from IPApp:

```
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 64,986B for [point, values, list, element] DOUBLE: 12,800 values, 102,912B raw, 64,939B comp, 1 pages, encodings: [RLE_PLAIN]
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 64B for [point, indices, list, element] INT32: 100 values, 14B raw, 30B comp, 1 pages, encodings: [RLE_PLAIN]
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 64,817B for [point, values, list, element] DOUBLE: 12,800 values, 102,912B raw, 64,779B comp, 1 pages, encodings: [RLE_PLAIN]
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 66,094B for [point, values, list, element] DOUBLE: 12,800 values, 102,912B raw, 66,047B comp, 1 pages, encodings: [RLE_PLAIN]
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 62B for [point, type] INT32: 100 values, 10B raw, 28B comp, 1 pages, encodings: [BIT_PACKED, PLAIN_DICTIONARY, RLE], d=1
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 50B for [point, type] INT32: 100 values, 7B raw, 27B comp, 1 pages, encodings: [BIT_PACKED, RLE, PLAIN]
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 53B for [point, indices, list, element] INT32: 100 values, 14B raw, 30B comp, 1 pages, encodings: [RLE, PLAIN]
16/04/06 17:57:26 INFO ColumnChunkPageWriteStore: written 65,899B for [point, values, list, element] DOUBLE: 12,800 values, 102,912B raw, 65,852B comp, 1 pages, encodings: [RLE]
16/04/06 17:57:26 INFO FileOutputCommitter: Saved output of task 'attempt_201604061757_0058_m_000002_0' to file:/C:/Users/npdar/Desktop/Acads/BigDataAnalytics/Assignment/Tutorial/parquet/part-r-00002
16/04/06 17:57:26 INFO FileOutputCommitter: Saved output of task 'attempt_201604061757_0058_m_000000_0' to file:/C:/Users/npdar/Desktop/Acads/BigDataAnalytics/Assignment/Tutorial/parquet/part-r-00000
16/04/06 17:57:26 INFO FileOutputCommitter: Saved output of task 'attempt_201604061757_0058_m_000001_0' to file:/C:/Users/npdar/Desktop/Acads/BigDataAnalytics/Assignment/Tutorial/parquet/part-r-00001
16/04/06 17:57:26 INFO FileOutputCommitter: Saved output of task 'attempt_201604061757_0058_m_000003_0' to file:/C:/Users/npdar/Desktop/Acads/BigDataAnalytics/Assignment/Tutorial/parquet/part-r-00003
16/04/06 17:57:26 INFO ParquetFileReader: Initiating action with parallelism: 5
Save Clusters to data3/model/clusters
16/04/06 17:57:26 INFO FileInputFormat: Total input paths to process : 1
16/04/06 17:57:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:57:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:57:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:57:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:57:27 INFO deprecation: mapred.min.split.size is deprecated. Instead, use mapreduce.input.fileinputformat.split.minsize
16/04/06 17:57:27 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskK
16/04/06 17:57:27 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskK
16/04/06 17:57:27 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskK
16/04/06 17:57:27 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskK
16/04/06 17:57:28 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 100 records.
16/04/06 17:57:28 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 100 records.
16/04/06 17:57:28 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 100 records.
16/04/06 17:57:28 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 100 records.
```

Histogram generation based on size specified:

The screenshot shows the IntelliJ IDEA interface with the following details:

- Project Structure:** Shows a tree view of the project structure under 'Image_Classification'.
- Code Editor:** Displays the file 'IPApp.scala' containing Scala code for a machine learning application.
- Run Output:** Shows the terminal output of the application's execution, displaying multiple histograms being generated.
- Status Bar:** At the bottom, it says "Compilation completed successfully in 7s 29ms (20 minutes ago)".
- Bottom Icons:** Standard Windows-style icons for file operations like Find, Run, Terminal, and TODO.

Model has been created based on the Random Forest Classifier Algorithm

The screenshot shows the IntelliJ IDEA interface. The top navigation bar includes File, Edit, View, Navigate, Code, Analyze, Refactor, Build, Run, Tools, VCS, Window, Help. The title bar indicates the project is 'Image_Classification' and the current file is 'IPApp.scala'. A message in the top right says 'Platform and Plugin Updates' and 'IntelliJ IDEA is ready to update'. The left sidebar shows the project structure with 'Project' selected, displaying 'data3' and its subfolders: 'model', 'test', 'keys', 'phone', 'specs', 'watch', and 'train'. The main code editor window contains Scala code for a Random Forest classifier. The bottom run log window shows the command-line output of the application running, including logs from 'ColumnChunkPageWriteStore', 'FileOutputCommitter', and 'ParquetFileReader'. The bottom status bar shows the date and time as '4/6/2016 6:09 PM'.

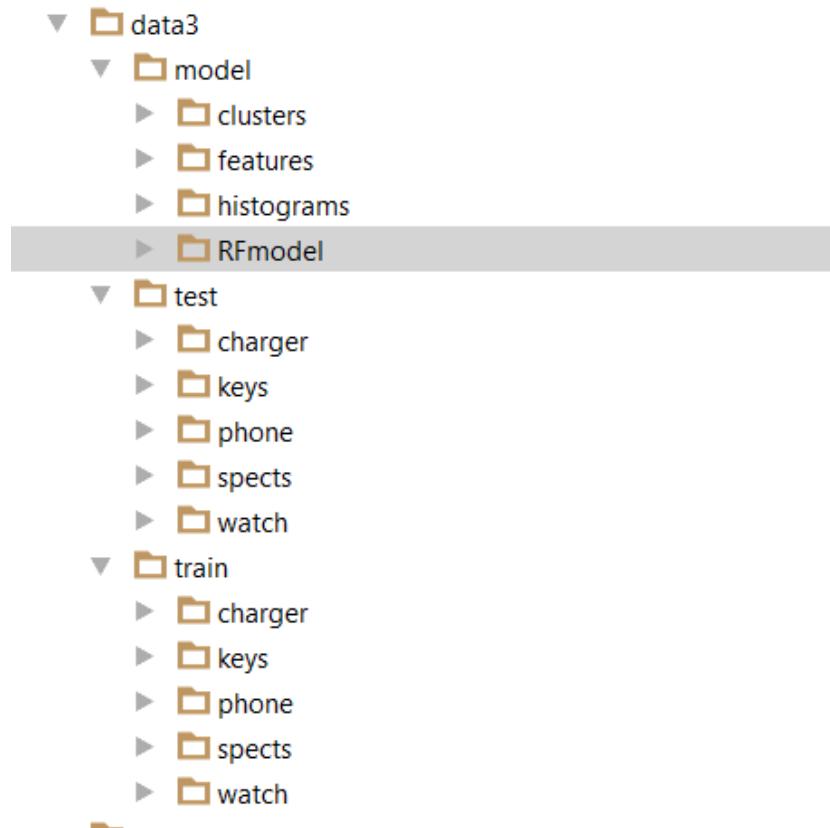
```

// Empty categoricalFeaturesInfo indicates all features are continuous.
val numClasses = 10
val categoricalFeaturesInfo = Map[Int, Int]()
// val numTrees = 10 // Use more in practice.
// val featureSubsetStrategy = "auto" // Let the algorithm choose.
// val impurity = "gini"
// val maxDepth = 4
val maxBins = 100

val numTrees = 4 to(5, 1)
val strategies = List("all", "sqrt", "log2", "onethird")
val maxDepths = 3 to(6, 1)
val impurities = List("gini", "entropy")

var bestModel: Option[RandomForestModel] = None
var bestErr = 1.0
val bestParams = new mutable.HashMap[Any, Any]()
var bestNumTrees = 0
var bestNumTreesSubSet = ""

```



Confusion Matrix:

Screenshot of IntelliJ IDEA 15.0.3 showing the code and output for generating a confusion matrix.

```

// Empty categoricalFeaturesInfo indicates all features are continuous.
val numClasses = 10
val categoricalFeaturesInfo = Map[Int, Int]()
// val numTrees = 10 // Use more in practice.
// val featureSubsetStrategy = "auto" // Let the algorithm choose.
// val impurity = "gini"
// val maxDepth = 4
val maxBins = 100

```

Output window shows the following data:

```

1.0 0.0 0.0 0.0
0.0 1.0 0.0 0.0
0.0 0.0 1.0 0.0
0.0 0.0 0.0 1.0
1.0
numTrees 5 featureSubsetStrategy onethird impurity entropy maxDepth 5
16/04/06 17:59:04 WARN DecisionTreeMetadata: DecisionTree reducing maxBins from 100 to 15 (= number of training instances)
Test Error = 0.0
=====
Confusion matrix =====
1.0 0.0 0.0
0.0 1.0 0.0
0.0 0.0 1.0
0.0 0.0 0.0 1.0
1.0
numTrees 5 featureSubsetStrategy onethird impurity entropy maxDepth 6
16/04/06 17:59:05 WARN DecisionTreeMetadata: DecisionTree reducing maxBins from 100 to 15 (= number of training instances)
Test Error = 0.25
=====
Confusion matrix =====
1.0 0.0 0.0
0.0 1.0 0.0
0.0 0.0 1.0
0.75
Best Err 0.0
Best params (featureSubsetStrategy,all) (numTrees,4) (maxDepth,3) (impurity,gini)
16/04/06 17:59:05 WARN DecisionTreeMetadata: DecisionTree reducing maxBins from 100 to 15 (= number of training instances)
16/04/06 17:59:06 INFO FileOutputCommitter: Saved output of task 'attempt_201604061759_1280_m_000000_4915' to file: C:/Users/npdar/Desktop/Acds/BigDataAnalytics/Assignment/Tut...
16/04/06 17:59:07 INFO CodecConfig: Compression: GZIP
16/04/06 17:59:07 INFO ParquetOutputFormat: Parquet block size to 134217728

```

Event Log shows 7754 LF+ UTF-8 messages at 6:10 PM on 4/6/2016.

Screenshot of IntelliJ IDEA 15.0.3 showing the code and output for generating a confusion matrix.

```

(0..,4)
(3,0,3)
(2,0,3)
(2,0,2)
(3,0,2)
(0,0,2)
(3,0,2)
(2,0,2)
(3,0,2)
(0,0,2)
(2,0,1)
(1,0,1)
(0,0,1)
(3,0,1)
(0,0,1)
(1,0,1)
(1,0,1)
(1,0,0)
(3,0,0)
(0,0,0)
(0,0,0)
(0,0,0)
(1,0,0)
(2,0,0)
0.3333333333333333
===== Confusion matrix =====
3.0 2.0 1.0 1.0 0.0
2.0 3.0 1.0 1.0 0.0
2.0 0.0 2.0 3.0 0.0
0.0 0.0 1.0 2.0 0.0
4.0 1.0 1.0 0.0 0.0
0.3333333333333333
16/04/06 18:00:51 INFO RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.
16/04/06 18:00:51 INFO RemoteActorRefProvider$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.

Process finished with exit code 0

```

Event Log shows 2577.1 LF+ UTF-8 messages at 6:12 PM on 4/6/2016.

Predicted Results Snapshots from the testing dataset

Screenshot of IntelliJ IDEA 15.0.3 showing the code editor and terminal for a Scala project named "Image_Classification".

The code in `IPApp.scala` contains logic for testing image classification:

```
// testImageClassification(sc)

val testImages = sc.wholeTextFiles(s"${IPSettings.INPUT_DIR}/*/*.jpg")
val testImagesArray = testImages.collect()
var predictionLabels = List[String]()
testImagesArray.foreach(f => {
    val splitStr = f._1.split("/")
    val predictedClass: Double = classifyImage(sc, splitStr(1))
    val segments = f._1.split("/")
    val cat = segments(segments.length - 2)
    val GivenClass = IMAGE_CATEGORIES.indexOf(cat)
    println(s"Predicting test image: " + cat + " as " + IMAGE_CATEGORIES(predictedClass.toInt))
    predictionLabels = predictedClass + ":" + GivenClass :: predictionLabels
})
```

The terminal shows the application's log output:

```
16/04/06 17:46:37 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 128
16/04/06 17:46:37 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 59 records.
16/04/06 17:46:37 INFO InternalParquetRecordReader: at row 0. reading next block
16/04/06 17:46:37 INFO CodecPool: Got brand-new decompressor [.gz]
16/04/06 17:46:37 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 59
16/04/06 17:46:37 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 138 records.
16/04/06 17:46:37 INFO InternalParquetRecordReader: at row 0. reading next block
16/04/06 17:46:37 INFO CodecPool: Got brand-new decompressor [.gz]
16/04/06 17:46:37 INFO InternalParquetRecordReader: block read in memory in 2 ms. row count = 138
Predicting test image : charger as charger
file:///C:/Users/npdar/Desktop/Acds/BigDataAnalytics/Assignment/Tutorial 9/CS5542 - Tutorial 9 Code/Image_Classification/data3/test/charger/charger4.jpg
16/04/06 17:46:37 INFO FileInputFormat: Total input paths to process : 1
16/04/06 17:46:37 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:46:37 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:46:37 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:46:37 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 17:46:38 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskAttempt
16/04/06 17:46:38 INFO ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskAttempt
16/04/06 17:46:38 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 100 records.
```

The status bar indicates "Compilation completed successfully in 7s 417ms (2 minutes ago)".

Screenshot of IntelliJ IDEA 15.0.3 showing the code editor and terminal for a Scala project named "Image_Classification".

The code in `IPApp.scala` contains logic for testing image classification:

```
// Empty categoricalFeaturesInfo indicates all features are continuous.
val numClasses = 10
val categoricalFeaturesInfo = Map[Int, Int]()
// val numTrees = 10 // Use more in practice.
// val featureSubsetStrategy = "auto" // Let the algorithm choose.
// val impurity = "gini"
// val maxDepth = 4
val maxBins = 100

val numOFTrees = 4 to(5, 1)
val strategies = List("all", "sqrt", "log2", "onethird")
val maxDepths = 3 to(6, 1)
val impurities = List("gini", "entropy")

var bestModel: Option[RandomForestModel] = None
var bestErr = 1.0
val bestParams = new mutable.HashMap[Any, Any]()
var bestnumTrees = 0
var bestFeatureSubSet = ""
```

The terminal shows the application's log output:

```
16/04/06 18:00:26 INFO InternalParquetRecordReader: at row 0. reading next block
16/04/06 18:00:26 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 7 records.
16/04/06 18:00:26 INFO InternalParquetRecordReader: at row 0. reading next block
16/04/06 18:00:26 INFO CodecPool: Got brand-new decompressor [.gz]
16/04/06 18:00:26 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 7
16/04/06 18:00:26 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 7
16/04/06 18:00:26 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 7 records.
16/04/06 18:00:26 INFO InternalParquetRecordReader: at row 0. reading next block
16/04/06 18:00:26 INFO CodecPool: Got brand-new decompressor [.gz]
16/04/06 18:00:26 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 7
Predicting test image : spectra as spectra
16/04/06 18:00:27 INFO FileInputFormat: Total input paths to process : 1
16/04/06 18:00:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 18:00:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 18:00:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 18:00:27 INFO ParquetFileReader: Initiating action with parallelism: 5
16/04/06 18:00:27 INFO ParquetFileReader: Initiating action with parallelism: 5
```

The status bar indicates "Compilation completed successfully in 7s 29ms (15 minutes ago)".

Image_Classification - [C:\Users\npdar\Desktop\Acads\BigDataAnalytics\Assignment\Assignment\Tutorial 9\CS5542 - Tutorial 9 Code]\Image_Classification] - [image_classification] - ..\src\main\scala\IPApp.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

Image_Classification > data3 > model >

Project Packages Project Files Problems

IPApp.scala x IPSettings.scala x build.sbt x

// Empty categoricalFeaturesInfo indicates all features are continuous.
 val numClasses = 10
 val categoricalFeaturesInfo = Map[Int, Int]()
 // val numTrees = 10 // Use more in practice.
 // val featureSubsetStrategy = "auto" // Let the algorithm choose.
 // val impurity = "gini"
 // val maxDepth = 4
 val maxBins = 100

val numOFTrees = 4 to(5, 1)
 val strategies = List("all", "sqrt", "log2", "onethird")
 val maxDepths = 3 to(6, 1)
 val impurities = List("gini", "entropy")

var bestModel: Option[RandomForestModel] = None
 var bestErr = 1.0
 val bestParams = new mutable.HashMap[Any, Any]()
 var bestnumTrees = 0
 var bestFeatureSubSet = ""

Run IPApp

16/04/06 18:00:18 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 7 records.
 16/04/06 18:00:18 INFO InternalParquetRecordReader: at row 0. reading next block
 16/04/06 18:00:18 INFO InternalParquetRecordReader: block read in memory in 0 ms. row count = 7
 16/04/06 18:00:18 INFO CodecPool: Got brand-new decompressor [.gz]
 16/04/06 18:00:18 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 7
 Predicting test image : phone as phone
 16/04/06 18:00:18 INFO FileInputFormat: Total input paths to process : 1
 16/04/06 18:00:18 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 18:00:18 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 18:00:18 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 18:00:18 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 18:00:18 INFO ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskA
 16/04/06 18:00:18 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskA
 16/04/06 18:00:18 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 100 records.
 16/04/06 18:00:18 INFO InternalParquetRecordReader: at row 0. reading next block
 16/04/06 18:00:18 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskA
 16/04/06 18:00:18 INFO CodecPool: Got brand-new decompressor [.gz]

Find Run Terminal TODO Event Log

Compilation completed successfully in 7s 29ms (16 minutes ago)

2153.72 LF+ UTF-8 605 PM 4/6/2016

Image_Classification - [C:\Users\npdar\Desktop\Acads\BigDataAnalytics\Assignment\Assignment\Tutorial 9\CS5542 - Tutorial 9 Code]\Image_Classification] - [image_classification] - ..\src\main\scala\IPApp.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

Image_Classification > data3 > model >

Project Packages Project Files Problems

IPApp.scala x IPSettings.scala x build.sbt x

// Empty categoricalFeaturesInfo indicates all features are continuous.
 val numClasses = 10
 val categoricalFeaturesInfo = Map[Int, Int]()
 // val numTrees = 10 // Use more in practice.
 // val featureSubsetStrategy = "auto" // Let the algorithm choose.
 // val impurity = "gini"
 // val maxDepth = 4
 val maxBins = 100

val numOFTrees = 4 to(5, 1)
 val strategies = List("all", "sqrt", "log2", "onethird")
 val maxDepths = 3 to(6, 1)
 val impurities = List("gini", "entropy")

var bestModel: Option[RandomForestModel] = None
 var bestErr = 1.0
 val bestParams = new mutable.HashMap[Any, Any]()
 var bestnumTrees = 0
 var bestFeatureSubSet = ""

Run IPApp

16/04/06 17:59:53 INFO CodecPool: Got brand-new decompressor [.gz]
 16/04/06 17:59:53 INFO InternalParquetRecordReader: block read in memory in 0 ms. row count = 7
 16/04/06 17:59:53 INFO CodecPool: Got brand-new decompressor [.gz]
 16/04/06 17:59:53 INFO InternalParquetRecordReader: block read in memory in 18 ms. row count = 7
 16/04/06 17:59:53 INFO ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskA
 16/04/06 17:59:53 INFO InternalParquetRecordReader: RecordReader initialized will read a total of 7 records.
 16/04/06 17:59:53 INFO CodecPool: Got brand-new decompressor [.gz]
 16/04/06 17:59:53 INFO InternalParquetRecordReader: block read in memory in 1 ms. row count = 7
 Predicting test image : keys as keys
 16/04/06 17:59:53 INFO FileInputFormat: Total input paths to process : 1
 16/04/06 17:59:53 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 17:59:53 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 17:59:53 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 17:59:53 INFO ParquetFileReader: Initiating action with parallelism: 5
 16/04/06 17:59:53 INFO ParquetRecordReader: Initiating action with parallelism: 5
 16/04/06 17:59:53 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskA
 16/04/06 17:59:53 WARN ParquetRecordReader: Can not initialize counter due to context is not a instance of TaskInputOutputContext, but is org.apache.hadoop.mapreduce.task.TaskA

Find Run Terminal TODO Event Log

Compilation completed successfully in 7s 29ms (17 minutes ago)

1765.55 LF+ UTF-8 607 PM 4/6/2016



47% 9:39 PM

Hello World!

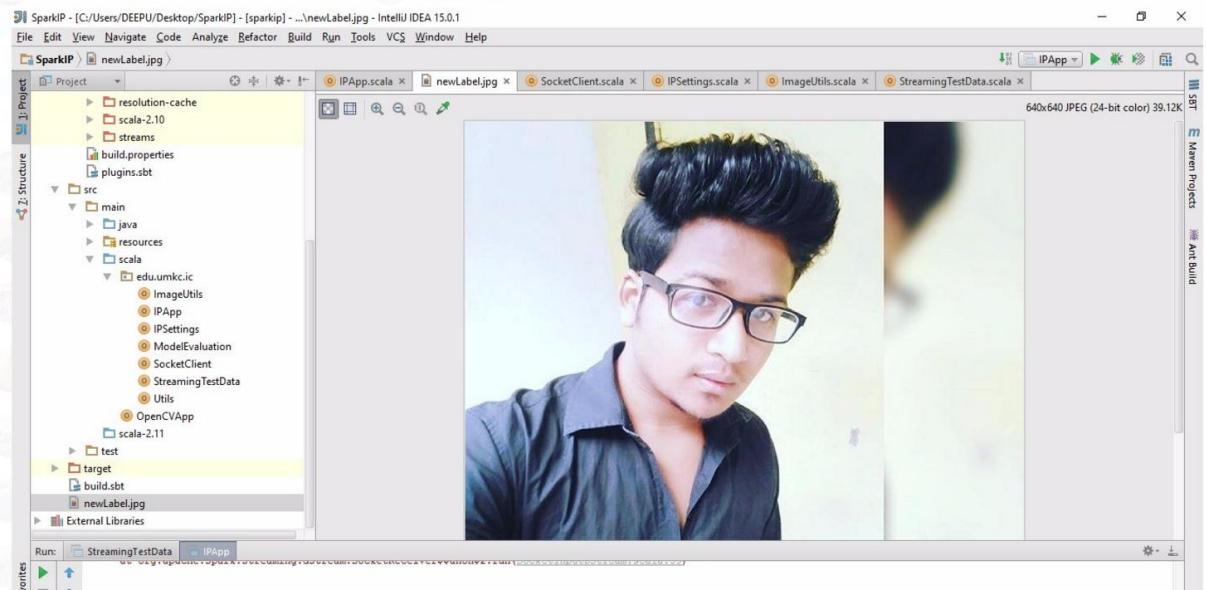
NOTIFY WEARABLE

I'm waiting here: 1234

SiteLocalAddress: 192.168.0.21

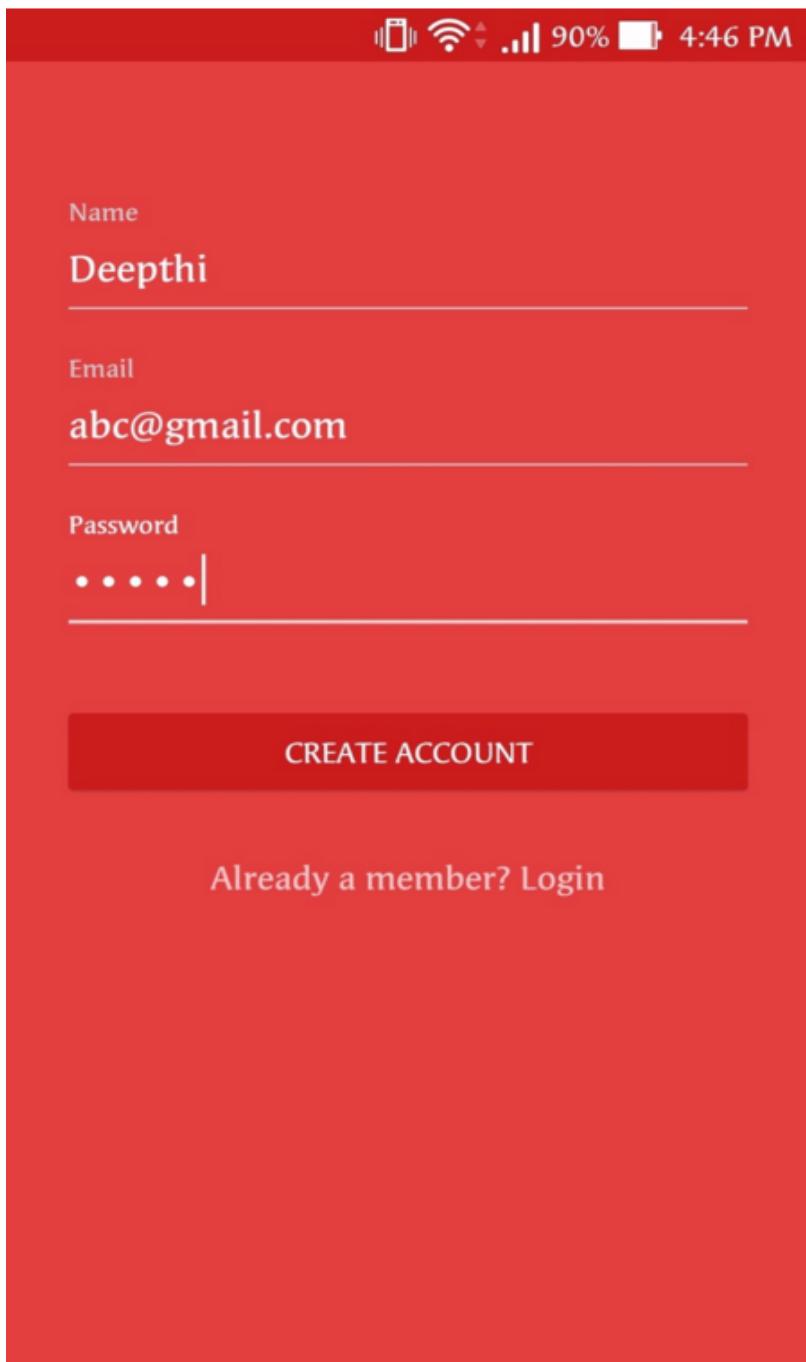
#1 from /192.168.0.23:53290

Predicted test image:spects/nreplayed: Hello from
Android, you are #1

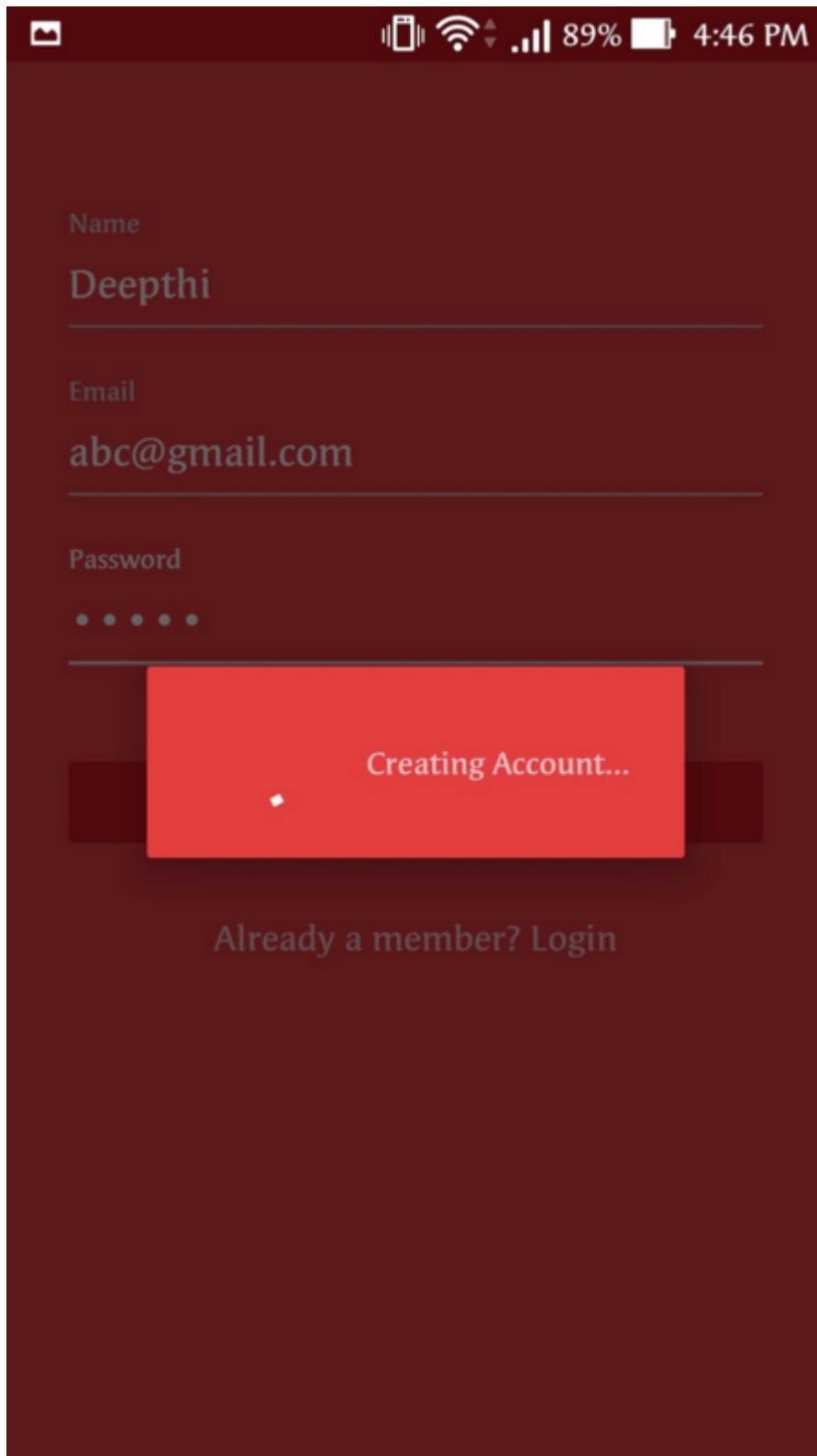


Final Phase Snapshots:

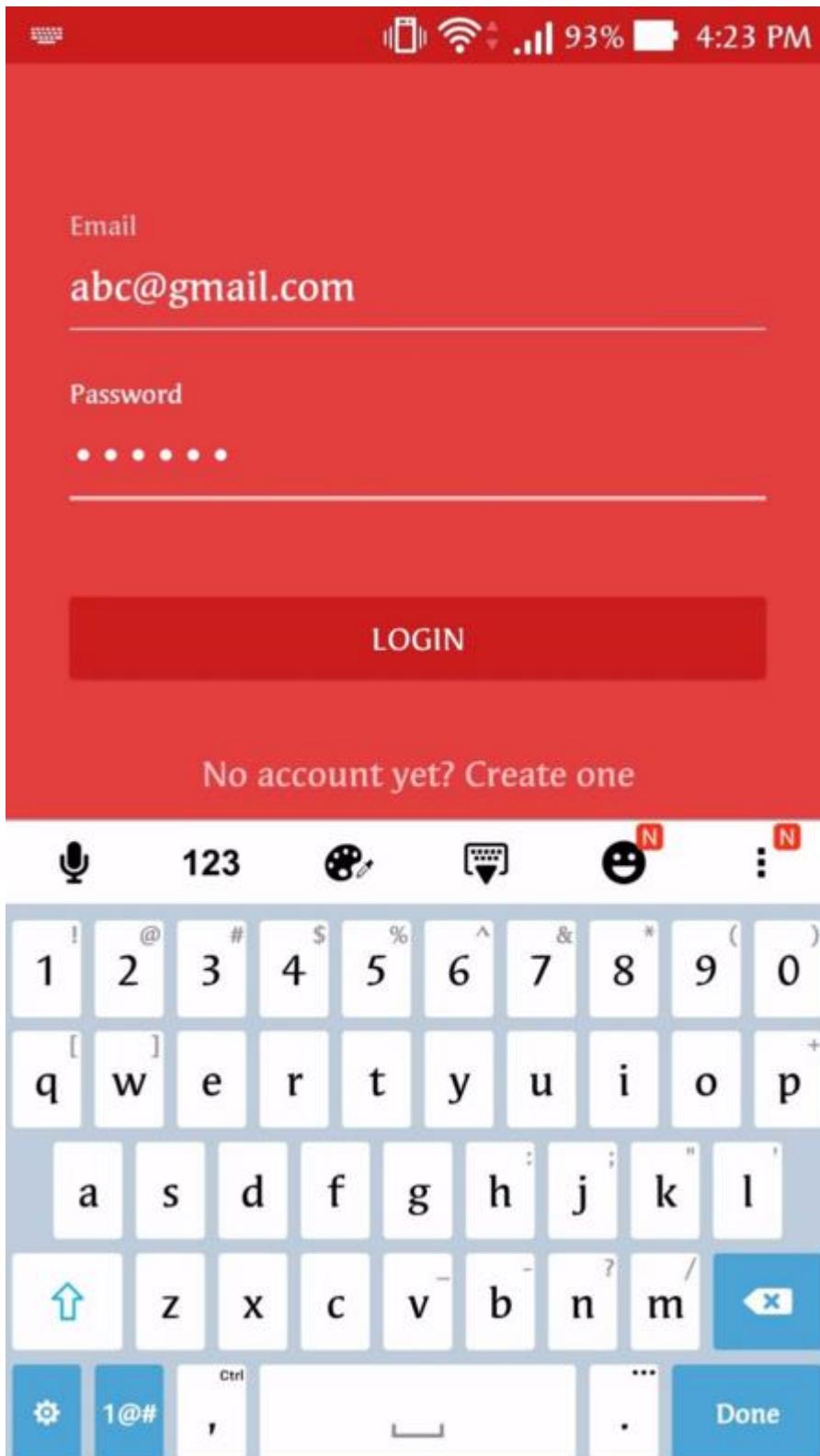
Android application - Registration page



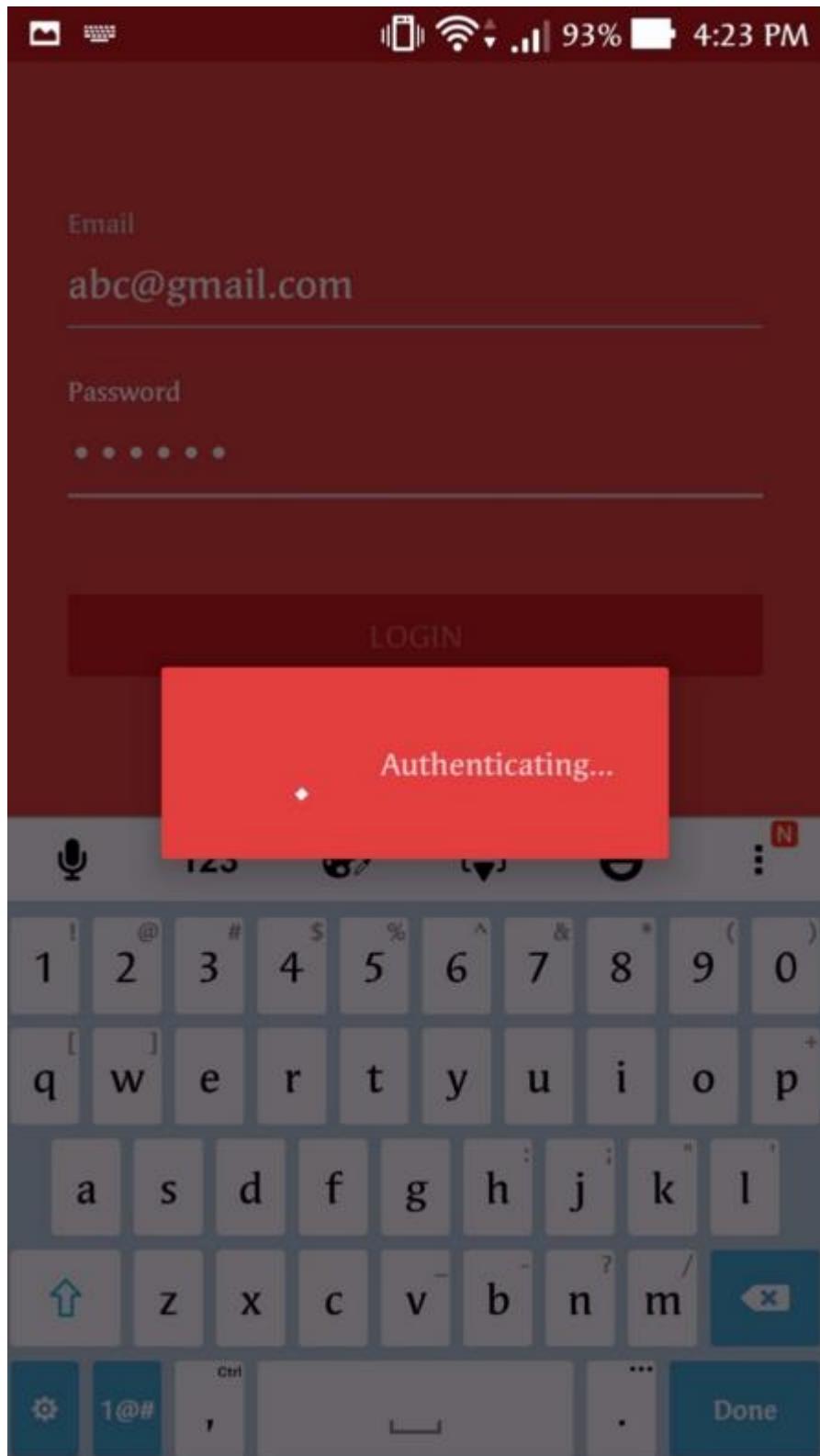
Snapshot to show that the account is being creating.



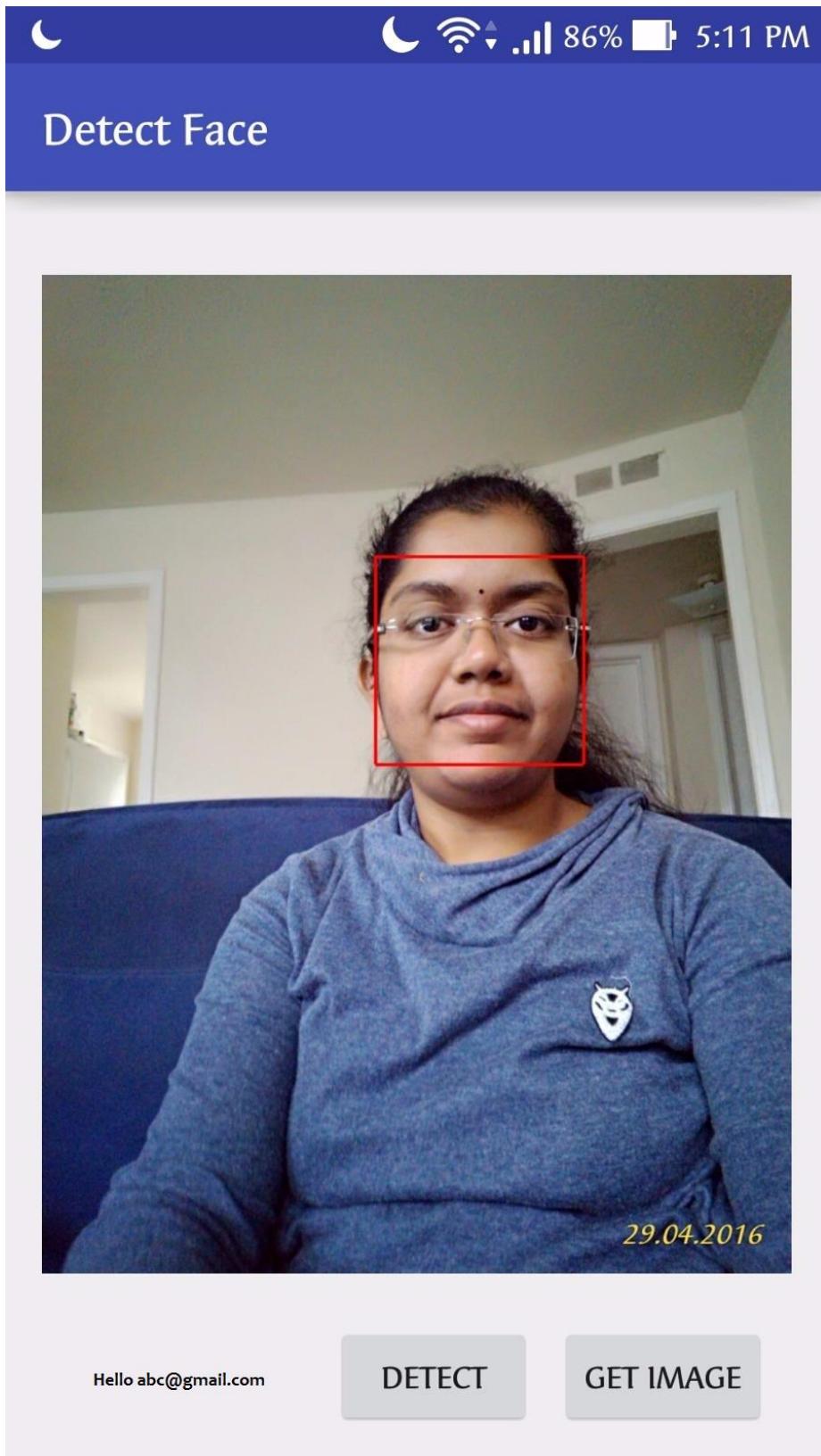
User login page after registration.



Account authentication Snapshot



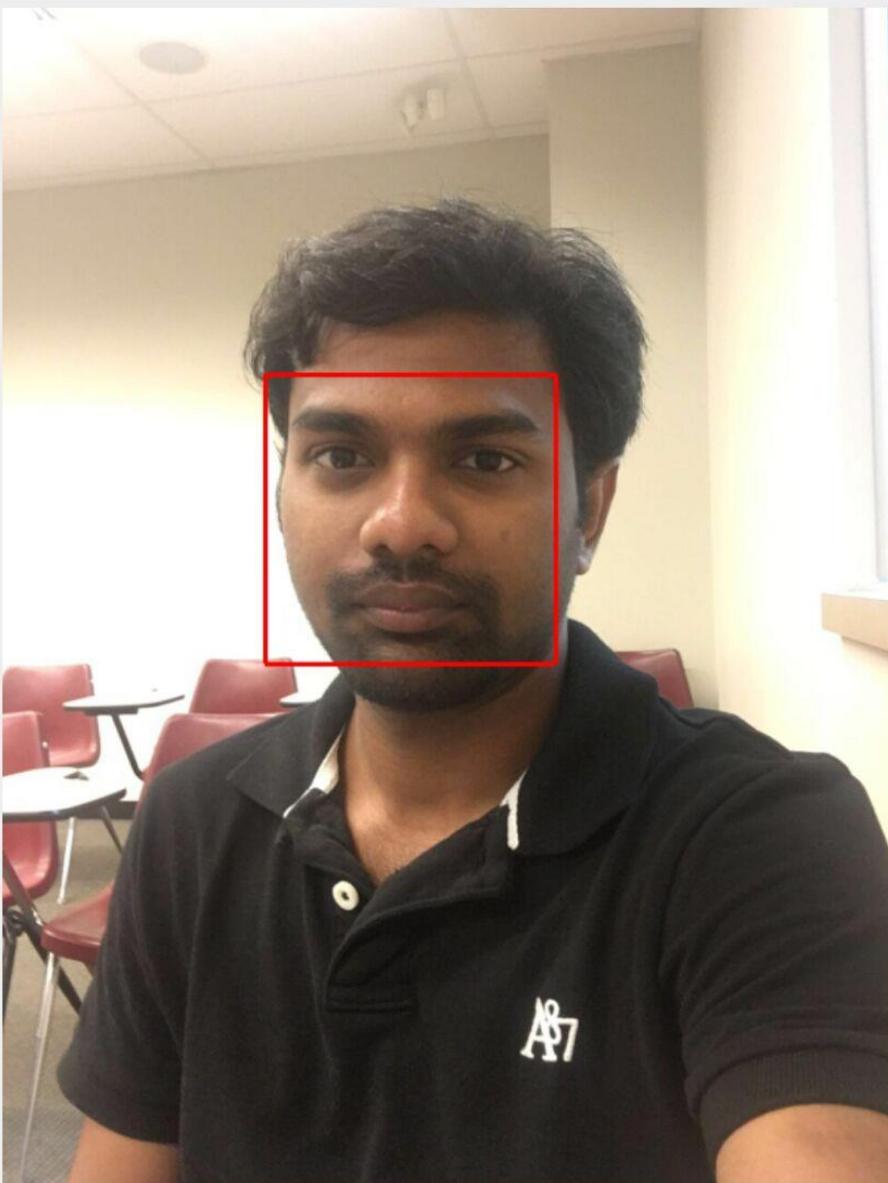
Facial Recognition





78% 8:34 PM

Detect Face

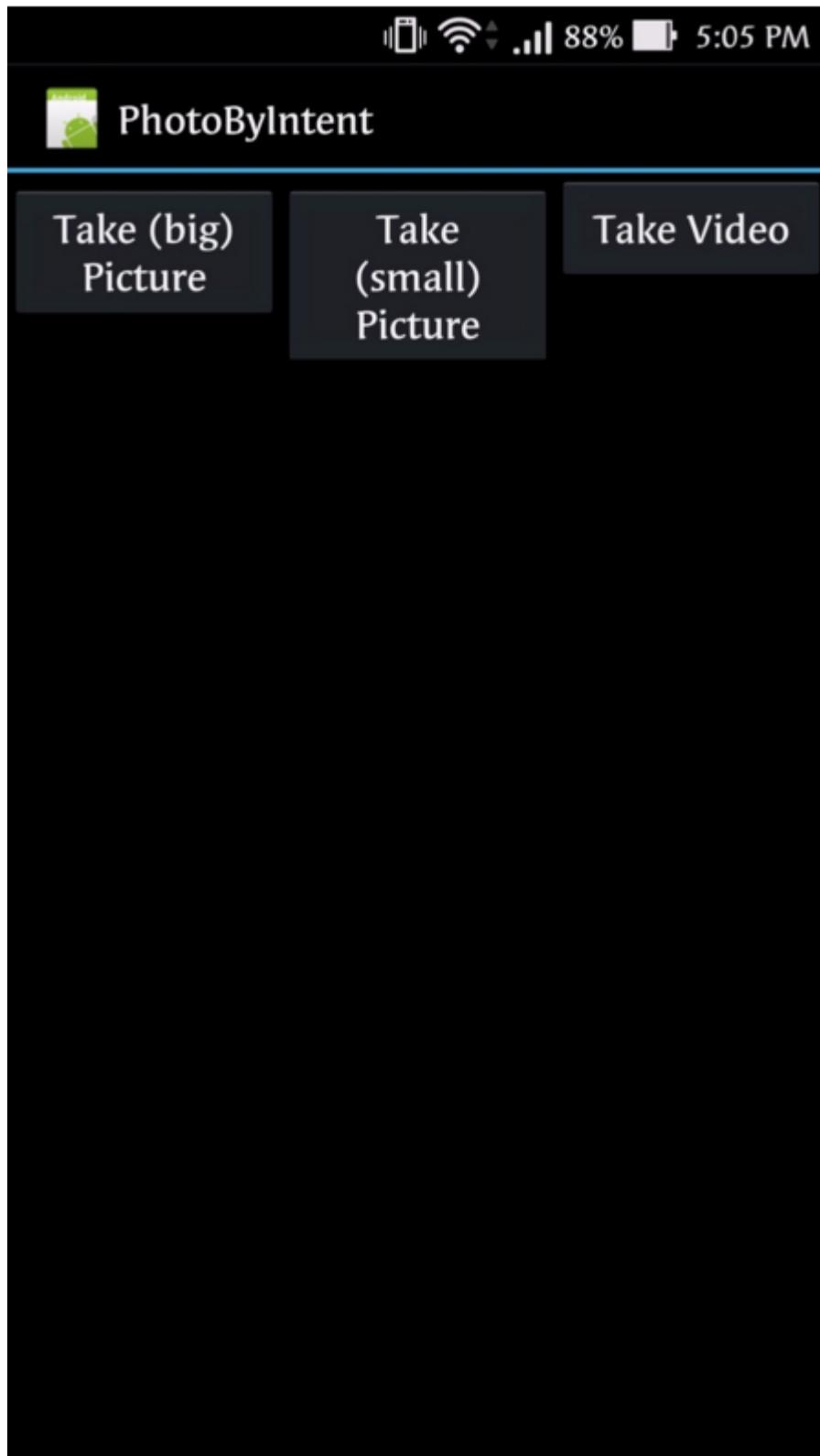


Hello tej@gmail.com

DETECT

GET IMAGE

Option to take an image





78% 8:37 PM



PhotoByIntent

Take (big)
Picture

Take
(small)
Picture

Take Video



Snapshot shows that the image has been saved into the Spark directory after receiving from Android client.

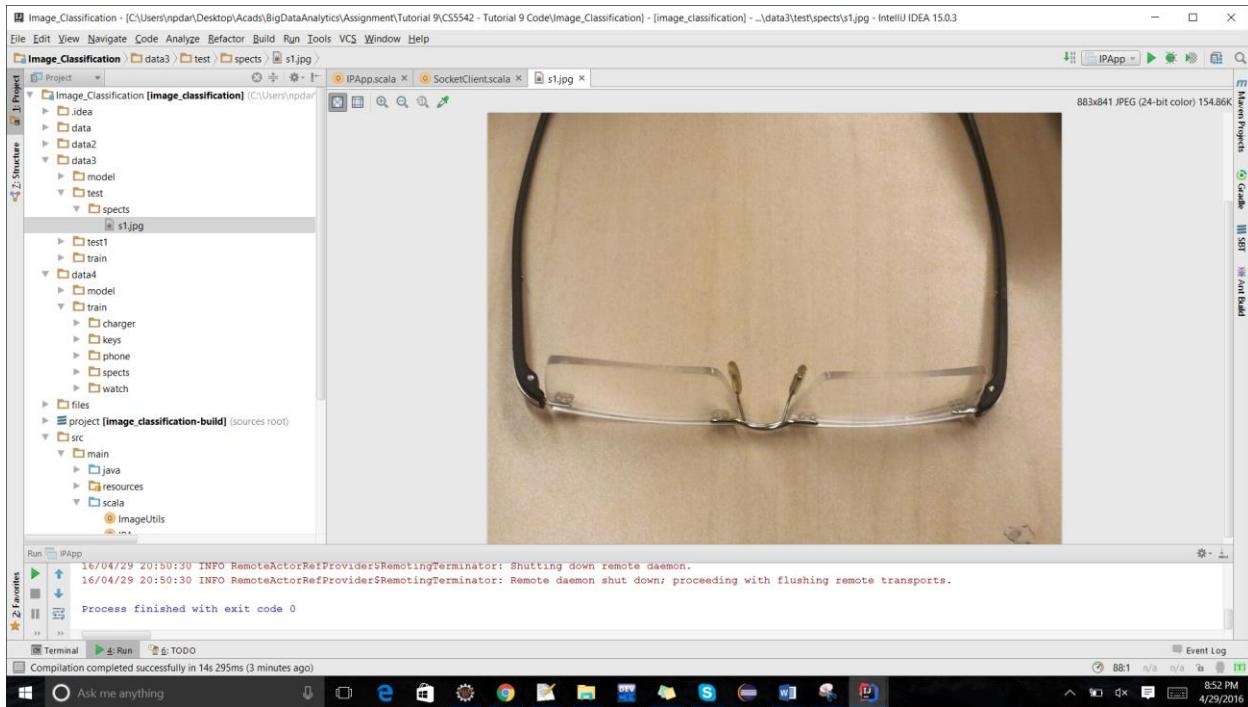
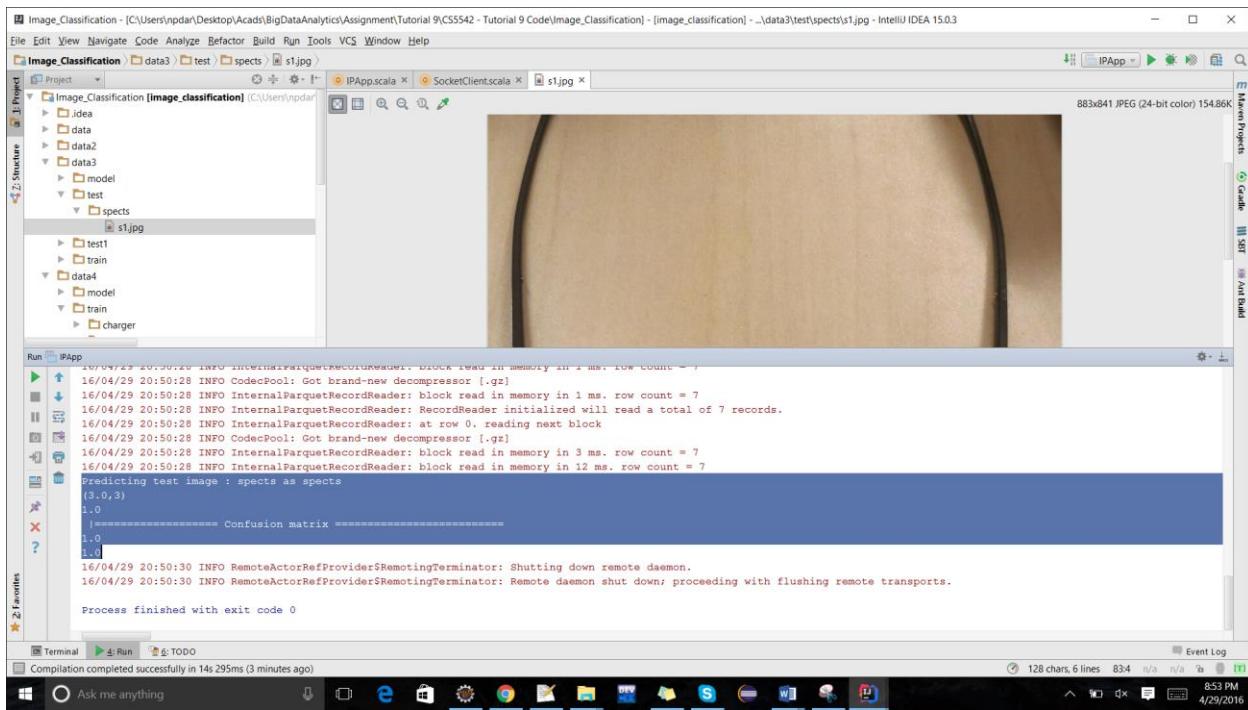
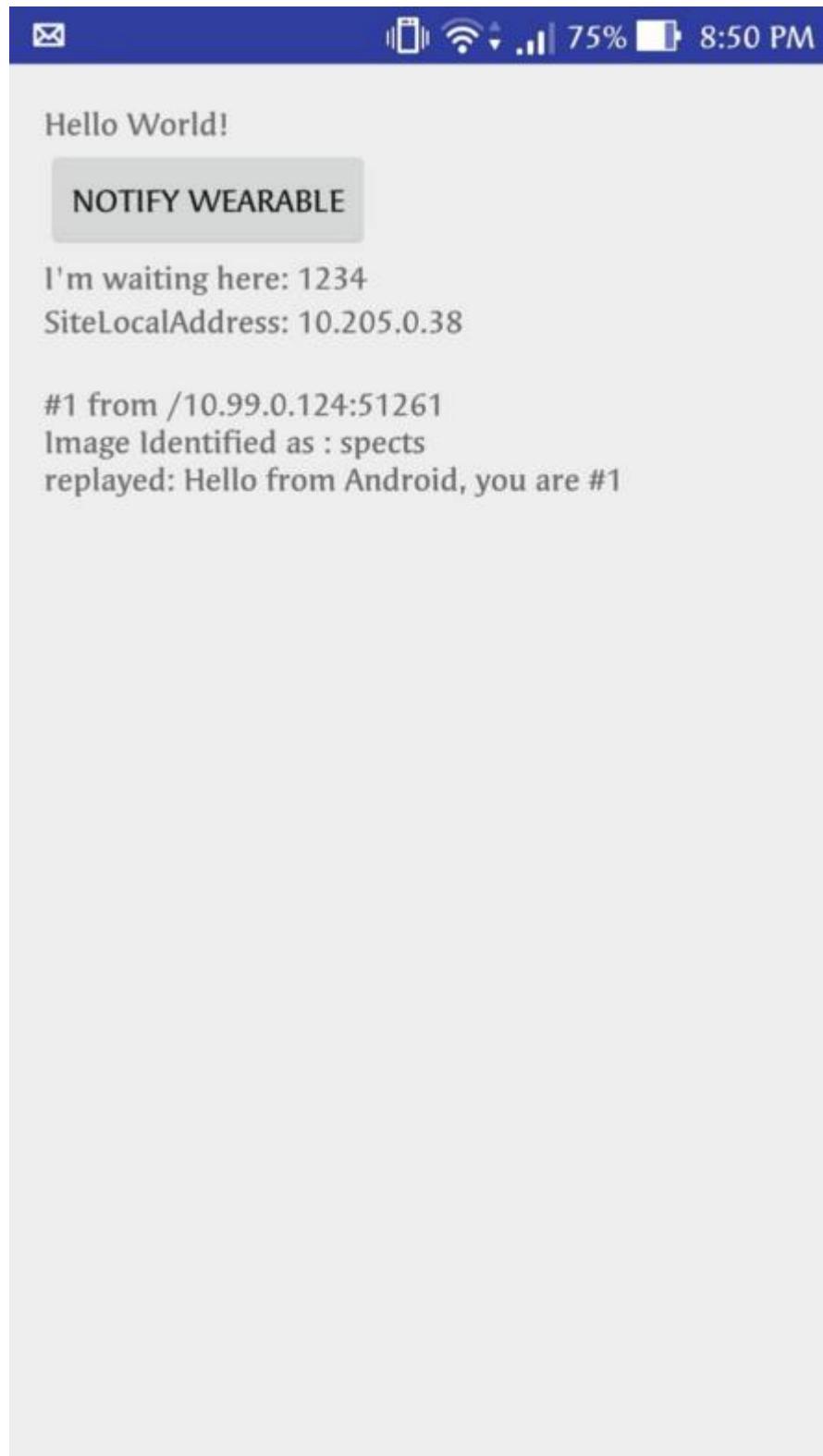


Image classification – Object classified as Specs



Notification to smart phone.



Recommendation system – To recommend the location of the object classified.

The screenshot shows the IntelliJ IDEA interface with the following details:

- Project Structure:** The project is named "SparkMachineLearnin". It contains a "movieLens" folder with "ObjectLocation", "location.dat", "ratings.dat", and "users.dat". A "src" folder contains "main" and "java" subfolders. "SparkCollaborativeFiltering.scala" is the active file.
- Code:** The code defines a main method for "SparkCollaborativeFiltering" that reads data from "test.data.txt", creates a SparkConf, and sets up a SparkContext. It then uses the ALS algorithm to build a recommendation model and prints the RMSE for validation datasets across different ranks (6, 8, 10, 12) and lambda values (0.1, 10.0, 20.0). The output shows that the best model has an RMSE of 3.7558695311242833 for rank=12, lambda=0.1, and numIter=20.
- Output:** The terminal pane shows the execution results, including the best model's details and a list of recommended locations: 1: on the table in living room, 2: in the cupboard @ bedroom, 3: underneath bed, 4: in the kitchen, 5: backyard garden.
- System Information:** The bottom right corner shows the date (4/29/2016), time (9:01 PM), and system status (Windows 10, ENG US).

Notification to Smart phone about recommendation



5.4 Deployment:

Git Hub Link:

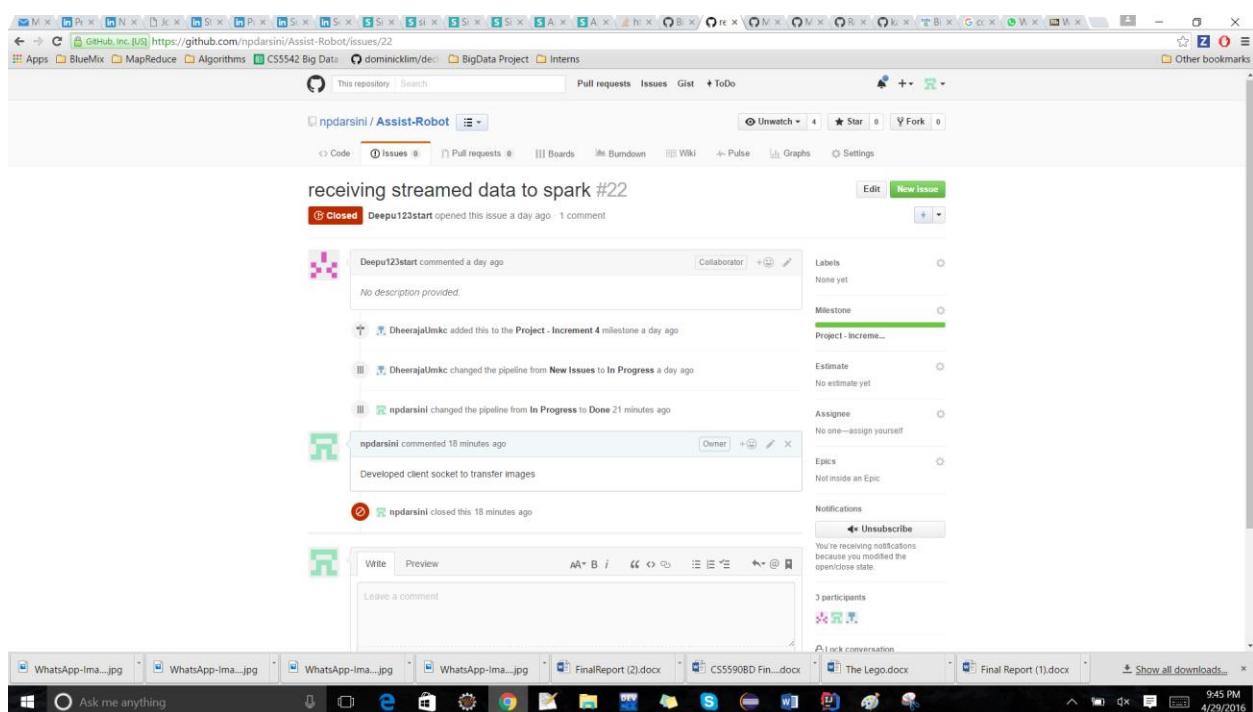
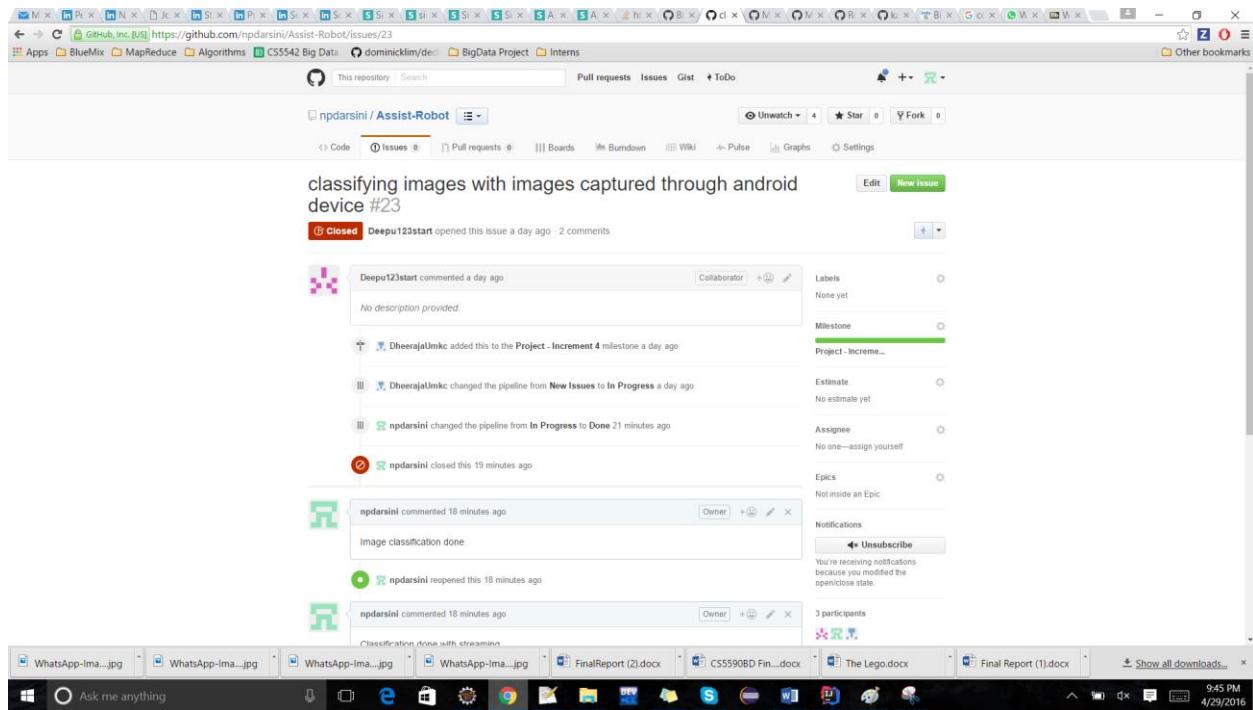
<https://github.com/npdarsini/Assist-Robot>

6. Project Management:

6.1 Implementation status report:

This phase involves the development of Image classification report with the streaming data. Basically the system will classify the object based on the provided image and will notify the user about the object. The image streaming is done from the android device for which we had created an application with the login option. This application can capture the image and transfers the image to the Spark IP address. Upon classifying the image the notification will be sent to the user device. Also a smart recommendation system has been developed which takes the classified image as a base input. Using this input data, it recommends the object location to the user. Assuming the user will be carrying the smart android device every single time, we thought that this could be as an innovative thought as the user don't need to spend a lot of time for finding the lost objects.

Name	Task	Contribution	Time Spent
Priyadarshini N	1. Data Streaming from Android to Spark through Socket communication. 2. Image Classification with streaming data from Android using Random Forest Algorithm technique.	100	25
Deepthi P	1. Data Streaming from Android to Spark through Socket communication. 2. Recommendation System for finding object location with ALS collaborative filtering technique.	100	25
Tej Kumar Y	1. Android camera module implementation to capture images. 2. Face recognition login.	100	20
Dheeraja V	1. Android login page 2. Notification on android device.	100	15



GitHub, Inc. [US] https://github.com/npdarsini/Assist-Robot/issues/21

Apps BlueMix MapReduce Algorithms CS5542 Big Data dominicklim/de... BigData Project Interns Other bookmarks

This repository Search Pull requests Issues Gist + ToDo

npdarsini / Assist-Robot Issues Boards Wiki Pulse Graphs Settings

Streaming data from android #21

Closed Deepu123start opened this issue a day ago · 1 comment

Deepu123start commented a day ago
No description provided

DheerajUmikc changed the pipeline from New Issues to In Progress a day ago

DheerajUmikc added this to the Project - Increment 4 milestone a day ago

npdarsini changed the pipeline from In Progress to Done 21 minutes ago

npdarsini changed the pipeline from Done to To Do 21 minutes ago

npdarsini changed the pipeline from To Do to In Progress 21 minutes ago

npdarsini changed the pipeline from In Progress to Done 21 minutes ago

npdarsini commented 18 minutes ago
Achieved using client socket programming.

npdarsini closed this 18 minutes ago

Lock conversation Notifications

Unsubscribe You're receiving notifications because you modified the specifice state.

3 participants

https://github.com/npdarsini/Assist-Robot/pulse

WhatsApp-Ima...jpg WhatsApp-Ima...jpg WhatsApp-Ima...jpg WhatsApp-Ima...jpg FinalReport (2).docx CS5590BD Fin...docx The Lego.docx Final Report (1).docx Show all downloads...

Ask me anything 945 PM 4/29/2016

GitHub, Inc. [US] https://github.com/npdarsini/Assist-Robot/issues/20

Apps BlueMix MapReduce Algorithms CS5542 Big Data dominicklim/de... BigData Project Interns Other bookmarks

This repository Search Pull requests Issues Gist + ToDo

npdarsini / Assist-Robot Issues Boards Wiki Pulse Graphs Settings

Linking Modules #20

Closed npdarsini opened this issue 23 days ago · 0 comments

npdarsini commented 23 days ago
No description provided

npdarsini added the enhancement label 23 days ago

npdarsini set the estimate to 13 23 days ago

DheerajUmikc added this to the Project - Increment 4 milestone a day ago

npdarsini closed this 22 minutes ago

Write Preview AA^b i Leave a comment

Styling with Markdown is supported Reopen issue Comment

Attach files by dragging & dropping, selecting them, or pasting from the clipboard.

2 participants

WhatsApp-Ima...jpg WhatsApp-Ima...jpg WhatsApp-Ima...jpg WhatsApp-Ima...jpg WhatsApp-Ima...jpg FinalReport (2).docx CS5590BD Fin...docx The Lego.docx Final Report (1).docx Show all downloads...

Ask me anything 946 PM 4/29/2016

Classification based on the Streaming Data #19

npdarsini opened this issue 33 minutes ago 0 comments

npdarsini commented 33 minutes ago
No description provided

npdarsini added enhancement, help wanted labels 32 minutes ago

Write Preview AA B i Leave a comment
Attach files by dragging & dropping, selecting them, or pasting from the clipboard
Styling with Markdown is supported Close issue Comment

Pipeline In Progress

Labels enhancement Help wanted

Milestone No milestone

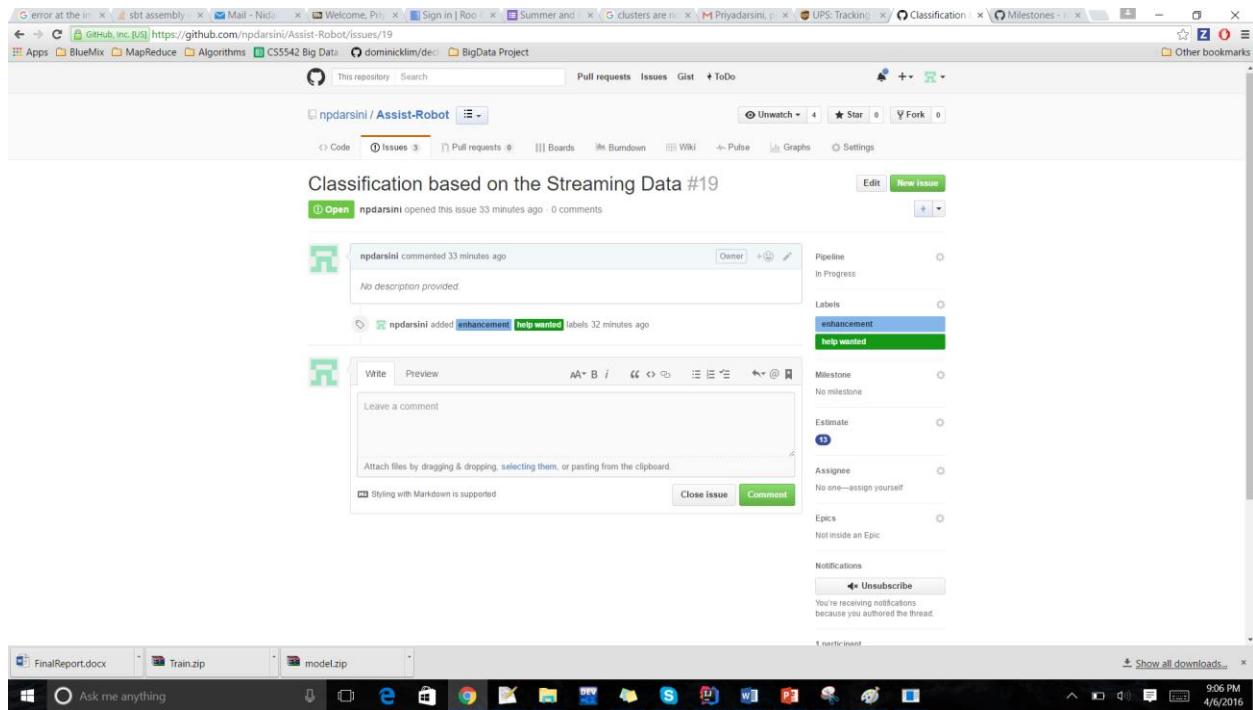
Estimate 13

Assignee No one—assign yourself

Epic Not inside an Epic

Notifications Unsubscribe You're receiving notifications because you authored the thread.

1 metric issue Show all downloads 9:06 PM 4/6/2016



Notification to Android Device - Recognized Object #18

DheerajaUmkc opened this issue a day ago 0 comments

DheerajaUmkc commented a day ago
To check if our project becomes better using Audio mining techniques
Not utilized

DheerajaUmkc added the enhancement label a day ago

DheerajaUmkc added this to the Project - Increment 3 milestone a day ago

DheerajaUmkc set the estimate to 5 a day ago

DheerajaUmkc closed this 7 hours ago

npdarsini changed the title from To utilize audio mining to Notification to Android Device - Recognized Object 35 minutes ago

Write Preview AA B i Leave a comment

Labels enhancement

Milestone Project - Increment 3

Estimate 5

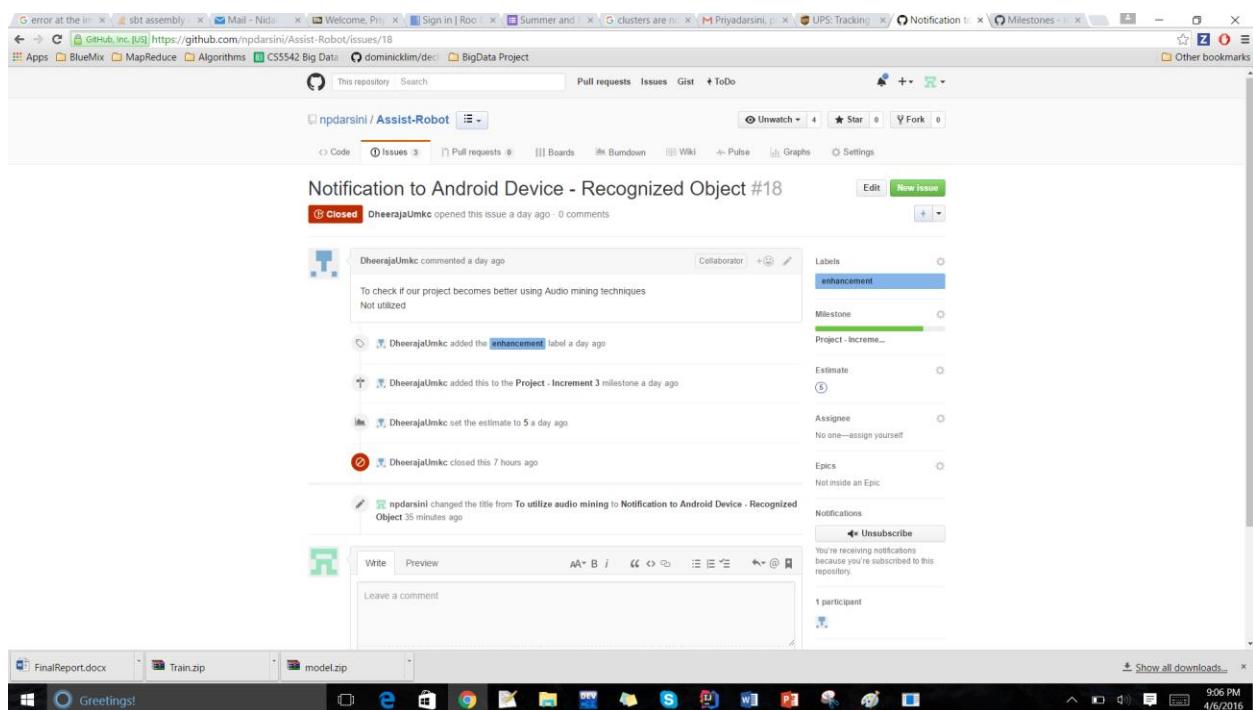
Assignee No one—assign yourself

Epic Not inside an Epic

Notifications Unsubscribe You're receiving notifications because you're subscribed to this repository.

1 participant Greetings!

FinalReport.docx Train.zip modelZip Show all downloads 9:06 PM 4/6/2016



Training data sets - Creation with different objects #17

Closed DheerajaUmkc opened this issue a day ago · 0 comments

DheerajaUmkc commented a day ago
Making use of text mining techniques

DheerajaUmkc added the **enhancement** label a day ago

DheerajaUmkc added this to the **Project - Increment 3** milestone a day ago

DheerajaUmkc set the estimate to **5** a day ago

DheerajaUmkc assigned Deepu123start and unassigned Deepu123start a day ago

DheerajaUmkc closed this a day ago

DheerajaUmkc reopened this a day ago

DheerajaUmkc closed this a day ago

npdarsini changed the title from **Usage of Text mining** to **Training data sets - Creation with different objects** 37 minutes ago

Labels: enhancement

Milestone: Project - Increment 3

Estimate: 5

Assignee: No one—assign yourself

Epic: Not inside an Epic

Notifications: Unsubscribe

2 participants

FinalReport.docx Train.zip modelZip Show all downloads 9:06 PM 4/6/2016

Exploring different Classification algorithms - Random Forest, Decision Tree #16

Closed DheerajaUmkc opened this issue 7 days ago · 0 comments

DheerajaUmkc commented 7 days ago
Trying to make use of classification algorithms and figure out the best that suits our application

Deepu123start was assigned by DheerajaUmkc 7 days ago

DheerajaUmkc added the **enhancement** label 7 days ago

DheerajaUmkc added this to the **Project - Increment 3** milestone 7 days ago

DheerajaUmkc set the estimate to **8** 7 days ago

DheerajaUmkc changed the estimate from **8** to **13** 7 days ago

DheerajaUmkc closed this a day ago

npdarsini changed the title from **Exploring different supervised learning algorithms** to **Exploring different Classification algorithms - Random Forest, Decision Tree** 38 minutes ago

Labels: enhancement

Milestone: Project - Increment 3

Estimate: 13

Assignee: Deepu123start

Epic: Not inside an Epic

Notifications: Unsubscribe

2 participants

FinalReport.docx Train.zip modelZip Show all downloads 9:05 PM 4/6/2016

GitHub Issues - npdarsini / Assist-Robot

Image classification #15

Closed DheerajaUmkc opened this issue 7 days ago · 0 comments

DheerajaUmkc commented 7 days ago
Image classification plays a vital role in our project

DheerajaUmkc self-assigned this 7 days ago

DheerajaUmkc added the **enhancement** label 7 days ago

DheerajaUmkc added this to the **Project - Increment 3** milestone 7 days ago

DheerajaUmkc set the estimate to 13 7 days ago

DheerajaUmkc changed the pipeline from **In Progress** to **Done** a day ago

DheerajaUmkc closed this a day ago

Labels: enhancement

Milestone: Project - Increment 3

Estimate: 13

Assignee: DheerajaUmkc

Epic: Not inside an Epic

Notifications: Unsubscribe

1 participant

Leave a comment

FinalReport.docx Train.zip model.zip

Ask me anything

9:05 PM 4/6/2016

GitHub Issues - npdarsini / Assist-Robot

Creating a perfect training Dataset and collecting Test data #11

Closed DheerajaUmkc opened this issue 4 hours ago · 0 comments

DheerajaUmkc commented 4 hours ago
Test data helps in finding the results and to understand the need for enhancement of the training data

DheerajaUmkc added the **enhancement** label 4 hours ago

DheerajaUmkc added this to the **Project - Increment 2** milestone 4 hours ago

DheerajaUmkc set the estimate to 3 4 hours ago

DheerajaUmkc changed the pipeline from **In Progress** to **Done** 3 hours ago

DheerajaUmkc closed this 2 hours ago

Labels: enhancement

Milestone: Project - Increment 2

Estimate: 3

Assignee: No one—assign yourself

Notifications: Unsubscribe

1 participant

Leave a comment

Attach files by dragging & dropping, selecting them, or pasting from the clipboard.

Comment Reopen issue Export issue

20160311182720.jpg Q4.jpg ProjectIncrem...docx Img4.jpg Img3.jpg Img2.jpg Img1.jpg CS5542-Lab7.pdf

Ask me anything

8:47 PM 3/11/2016

GitHub Issues: Data Collection and analysis #9

Closed DheerajUmkc opened this issue 4 days ago · 0 comments

Labels: enhancement

Milestone: Project - Increment 2

Estimate: 3 hours

Assignee: npdarsini

Notifications: Unsubscribe

Participants: 2

Lock conversation

More Issue

Export Issue

Comments:

- DheerajUmkc commented 4 days ago: Collection of data related to project, analyzing it and making it useful to the project.
- DheerajUmkc added this to the Project - Increment 2 milestone 4 days ago.
- DheerajUmkc added the enhancement label 4 days ago.
- DheerajUmkc set the estimate to 3 hours ago.
- DheerajUmkc changed the pipeline from New Issues to In Progress 4 hours ago.
- DheerajUmkc changed the estimate from 3 to 5 hours ago.
- npdarsini was assigned by DheerajUmkc 4 hours ago.
- DheerajUmkc closed this 4 hours ago.
- DheerajUmkc reopened this 4 hours ago.
- DheerajUmkc changed the title from Issue#9 to Data Collection and analysis 4 hours ago.
- DheerajUmkc changed the estimate from 5 to 3 hours ago.
- DheerajUmkc changed the estimate from 3 to 5 hours ago.
- DheerajUmkc changed the estimate from 5 to 3 hours ago.

Downloads: 20160311182720.jpg, Q4.jpg, ProjectIncrem...docx, Img4.jpg, Img3.jpg, Img2.jpg, Img1.jpg, CS5542-Lab7.pdf

Windows Taskbar: Ask me anything, 8:46 PM, 3/11/2016

GitHub Issues: Establishing connection between smart watch and Robome #8

Closed DheerajUmkc opened this issue 4 days ago · 0 comments

Labels: enhancement

Milestone: Project - Increment 2

Estimate: 3 hours

Assignee: npdarsini

Notifications: Unsubscribe

Participants: 2

Lock conversation

More Issue

Export Issue

Comments:

- DheerajUmkc commented 4 days ago: We send a notification from robome to smartwatch.
- DheerajUmkc added the enhancement label 4 days ago.
- npdarsini was assigned by DheerajUmkc 4 days ago.
- DheerajUmkc set the estimate to 3 hours ago.
- npdarsini changed the pipeline from To Do to Backlog 3 days ago.
- DheerajUmkc changed the pipeline from Backlog to Done 5 hours ago.
- DheerajUmkc added this to the Project - Increment 2 milestone 4 days ago.
- DheerajUmkc changed the title from Issue#8 to Establishing connection between smart watch and Robome 4 hours ago.
- DheerajUmkc closed this 4 hours ago.
- DheerajUmkc reopened this 4 hours ago.
- DheerajUmkc closed this 3 hours ago.

Downloads: 20160311182720.jpg, Q4.jpg, ProjectIncrem...docx, Img4.jpg, Img3.jpg, Img2.jpg, Img1.jpg, CS5542-Lab7.pdf

Windows Taskbar: Ask me anything, 8:45 PM, 3/11/2016

GitHub Issues - Installation of Spark #7

Closed DheerajUmkc opened this issue 4 days ago · 0 comments

DheerajUmkc commented 4 days ago Since usage of spark makes the project more flexible because of availability of RDD's

DheerajUmkc added the enhancement label 4 days ago

dunil210 was assigned by DheerajUmkc 4 days ago

DheerajUmkc added this to the Project - Increment 2 milestone 4 days ago

DheerajUmkc set the estimate to 3-4 days ago

npdarsini changed the pipeline from To Do to Backlog 3 days ago

DheerajUmkc changed the title from Issue3 to Installation of Spark 4 hours ago

dunil210 was unassigned by DheerajUmkc 4 hours ago

DheerajUmkc changed the pipeline from Backlog to Done 4 hours ago

Deepur123start was assigned by DheerajUmkc 4 hours ago

DheerajUmkc closed this 4 hours ago

DheerajUmkc reopened this 4 hours ago

DheerajUmkc closed this 4 hours ago

Ask me anything

GitHub Issues - Features of RoboMe #6

Closed DheerajUmkc opened this issue 4 days ago · 0 comments

DheerajUmkc commented 4 days ago Going through the features of Robome and what all can be derived using its basic features

DheerajUmkc added enhancement question labels 4 days ago

Deepur123start was assigned by DheerajUmkc 4 days ago

DheerajUmkc added this to the Project - Increment 2 milestone 4 days ago

DheerajUmkc set the estimate to 5-4 days ago

npdarsini changed the pipeline from In Progress to To Do 3 days ago

npdarsini changed the pipeline from To Do to In Progress 3 days ago

DheerajUmkc changed the pipeline from In Progress to Done 5 hours ago

DheerajUmkc removed the enhancement label 4 hours ago

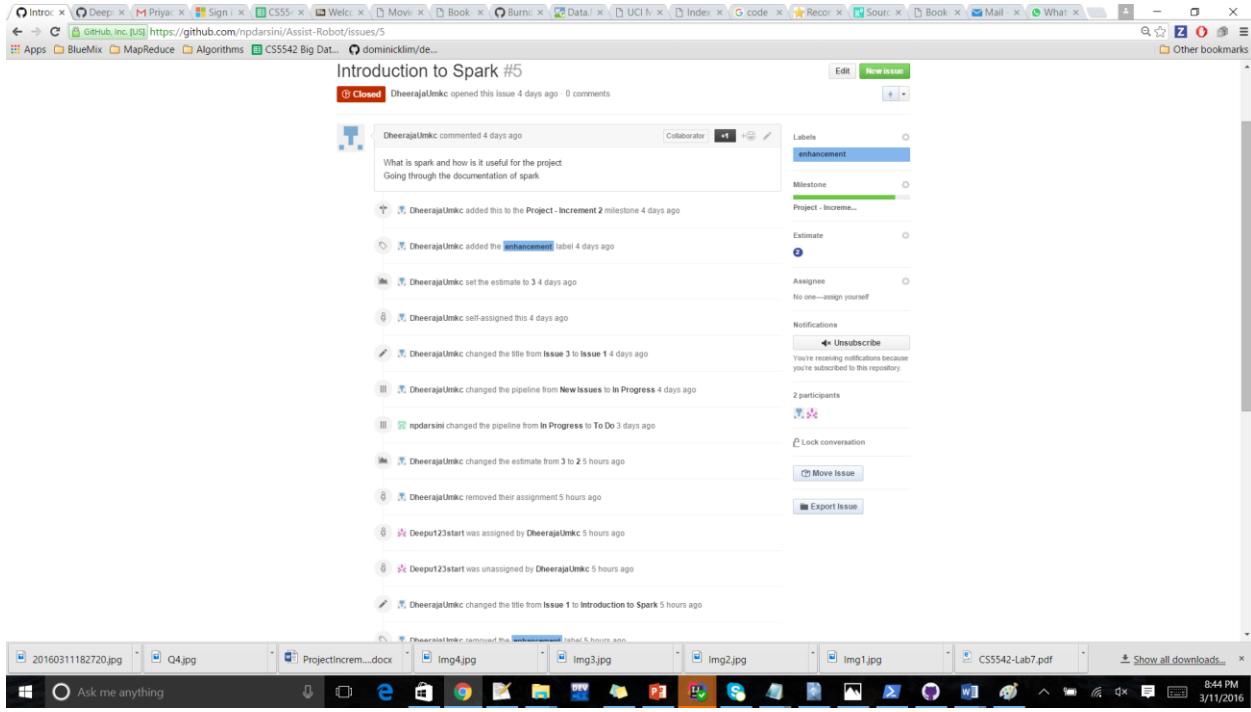
Deepur123start was unassigned by DheerajUmkc 4 hours ago

DheerajUmkc changed the estimate from 5 to 4 hours ago

npdarsini was assigned by DheerajUmkc 4 hours ago

npdarsini was unassigned by DheerajUmkc 4 hours ago

Ask me anything



7. Future Scope:

To implement a Robo based system that can move around and track the objects. Upon tracking, the user will be notified about the object location. To explore SparkSQL to store the captured images so that we will be able to deal with the massive data. To explore much about the Machine learning algorithms so as to achieve better accuracy in Image Classification. To embed Natural Language Processing techniques into the current working project.

8. Bibliography:

Lab Tutorials and the material provided by Dr. Lee.

SPARK - <https://spark.apache.org/docs/latest/programming-guide.html#external-datasets>

Spark Recommendation - <http://spark.apache.org/docs/latest/mllib-collaborative-filtering.html>