



# CS 5542 Project Summary


Gharibi, Mohamed (7)

Huang, Wei (9)

Sattineni, Vipin Reddy (24)

Xia, Ting (30)

## Introduction



Nowadays, the technology is getting improved day after day and hour after hour. From this point, all the people start looking for new ideas and projects which help to simplify the daily issues for the people. Moreover, as we know the data in different types through different social media websites and applications are getting improved and increased every single second as well. Especially, when we talk about large amount of data and how to use this data in an efficient way to serve us better.

There are many majors and many different ways to get use of this data so that it would be sufficient for all the people in an easy, fast and simple way. For example, having an application that matches the interests of people and makes recommendations based on the common interests can help in people bonding together stronger.

In our project we have created an application which collect the pictures from Instagram and perform the sentiment analysis and run the machine learning algorithms on these collected pictures over Spark. Moreover, after we have done this we make the recommendations based on our results. After that, we get other pictures and compare it with our results to check whether the image category and the recommendations are the same or not. Moreover, we are collecting the streaming tweets form Twitter and make the sentiment analysis over these tweets. As the previous part, the main point here is to make the recommendation after having the sentiment analysis results. Note: The main idea of the project is that making common interests of a user according to his activities' in Twitter and Instagram.

## Architecture & Applications

It contains three devices: the android smart watch, android phone, iPhone. The user can control the android phone and smartwatch to communicate with the iPhone. Also, iPhone can take photo when required. The photo will be analyzed and the results will be sent to the user. Spark is used to analyze the data sets and Mongolab is used to store the data. The smartwatch and phone are paired, so the communication is easy between the two.



### Machine Learning Algorithms:

In Hadoop/Machine learning, we are planning to use it as one of our core techniques to processing the data. In the later on project increment, it will behave as the part that “forces” the robot to react upon certain data we have collected or we have trained to the data. For example, in SparkR, we can use it to process certain amount of data to calculate K-Means, K centroid and Expectation Maximization(EM) and use it as a best representation of certain amount of data which fall under a defined category. Consequently, we can shrink the static data size into smaller ones so that it is easier to process. Specifically, the following services are used in our project.

#### a. **Use Spark MLlib:**

1. K-means analysis of heart rate.
2. Categorize collected tweets using Feature Extraction – TF-IDF.

a. **Sentiment analysis:** Stanford Core NLP is used to do sentiment analysis of collected tweets

b. **Recommendation:** Alternating Least Squares (ALS) is used to recommendation to the user.

### Datasets:

- a. Twitter streaming to collect tweets
- b. Sensor data get from Smartphone/smart watch

### Implementation – Image Recognition:

We used Alchemy API to process the pictures and when a user chooses an image then click "classify image" Alchemy API could recognize the picture and provide the tags for this picture (i.e. it was successfully recognize the cat image). Of course, as you can see in the picture above that this picture was not deployed on the computer, it was deployed on the Smartphone device. Just like that, we can implement this API on the smartwatch we have. Moreover, Alchemy API could help us in our project to classify the owner of the picture if he is a friend or not, since we are making our project as a home assistant application.

Recommendation to the User:

We analyzed the tweets and made recommendations, we made a sentiment analysis as well and able to push all the results to the android phone.

**Step 1:** Define 10 categories:

**Step 2:** Collect tweets as training data to categorize the tweets into these 10 categories using keywords searching.

**Step 3:** Collect tweets again for recommendation training. In order to do rating, four items should be collected:

- 1) UserId. It's the tweet's UserId which should be converted into integer.
- 2) Category. It is analyzed using feature extraction – TF-IDF by the training data collected in step 2.
- 3) Rating. Use sentiment analysis to give the rating for each tweet.
- 4) Timestamp. Tweet creation time which should be converted into integer.

These four items should be written into one file called "rating.txt".

**Step 4:** Get the category mapping file called "category.txt".

**Step 5:** Get the recommendation for one particular user. The recommendation is the categories that the user prefers.

**Step 6:** Send the results to the Smartphone/smart watch.

### Image Recommendation:

In our application we provided many services that could help a user a lot to improve his search and recommendation part. For example, some people like animals a lot so in this case we recommend animals pages and pictures for a user depends on the animal species he or she likes. Others, want to teach their children about animals and differences between them, so that we also implemented a service which allows you to recognize the animal species from many different categories. First of all, we implemented the text sentiment analysis which allows us to know what the user likes and hates, so that if the user typed something good about a cat, we know this user likes cats by using sentiment analysis.

In addition, through our application you will be able to choose an animal picture and process it using Spark. Our application will process the picture using Spark and will let you know weather this animal is a cat or dog, etc. We have ten animal categories to

recognize which animal you like. Then, we will recommend you pages and pictures for the same animal you like.

Through Spark, we will compare the picture you uploaded with another picture from Instagram and will let you know the opinion about this picture depending on the information we already had (Sentiment Analysis). Here is the main part of machine learning since every time we run another picture, our knowledge about this user increases. Also, the recommendation part will keep getting better and better.

In our application we were so strict for comparing the pictures in Spark so that the user will get the best result and later this result will be compared to the ten animal categories we used in our application. These categories contain animals' species such as [ dogs, cats, birds, etc.]. The main part is to define the picture for which category it belongs so that all the coming recommendations are related to this result.

**Step 1:** Define 9 categories:

**Step 2:** Collect images as training data to categorize the image into these 9 categories using image tag searching.

**Step 3:** Collect image from device

**Step 4:** Train the images in Spark.

**Step 5:** Categorize the image data.

**Step 6:** Do recommendation of the user according to sentimental analysis.

**Step 7:** Compare the image category collected from device and the user recommendation category.

**Step 8:** Send the comparison result to the device.

## Related Work

## Future Work

Our application mainly focuses on the common interests of the users on a broader level based on the user's activities on Instagram and Twitter. The future scope of our application would be to narrow down to a more specific interest level and also not only use the streaming data but use data from the past of the user i.e. use the previous posts or ratings of the users as the data to make the recommendations specific to a user.