2017 Spring: CS5542 Big Data Analytics and Apps
Technical Background & Project Development

Last Name: SURDARAJU

First Name: LAVA KUMAR

UMKC email: lsz85@mail.umkc.edu

- Machine Learning (Scale 0 – 10): 4
- Deep Learning (TensorFlow, Caffe, etc) (Scale 0 – 10): 4
- Spark MLlib (Scale 0 – 10): 6
- Hadoop (Scale 0 – 10): 6
- Cloudera (Scale 0 – 10): 5
- Android App (Scale 0 – 10): 8
- Java (Scale 0 – 10): 8
- Scala (Scale 0 – 10): 6
- Python (Scale 0 – 10): 6
- JavaScript (Scale 0 – 10): 8
- Projects: Have you ever worked on a software development project? If yes, describe them.
  1. Worked in a development project for an insurance client.
  2.
  3.

- Project Partners for Your CS5542 Project (if any) Independent Research - DR
  1. Last Name:
     First Name: N/A
  2. Last Name:
     First Name: N/A
  3. Last Name:
     First Name: N/A

Indicate if you have your own devices. If yes, describe about your device.

	Yes/No	System Spec (Processor, Memory, Hard Drive, Graphic Card, Operating System, etc)
Laptop	Yes	Dell - I5 - 1TB - 8GB - Windows 10
Smartphone	Yes	Iphone 6
Desktop	No	

CS5542 Big Data Analytics and Apps

Problem Set-1 (PS-1)

Jan. 17 (T), 2017

Name: LAVA KUMAR SURPARAJU

Email: lsz85@mail.umkc.edu

1. **Question: Odd Man Out**
   You're given an unsorted array of integers where every integer appears exactly twice, except for one integer which appears only once. Write an algorithm (in a language of your choice) that finds the integer that appears only once.
2. **System Design: Big Data Applications**
   A group of primatologists wants to study the details of the daily movement, activities, and interactions of a group of 6 chimpanzees living on "chimp island" - a natural, though somewhat open habitat about 50 meters in diameter, bounded on all sides by water, in the San-Diego zoo. Since they don't want to sit all day every day recording the second-by second positions and activities of the chimps, they have come to you, a computer vision expert, for automated assistance. They are interested in both compiling statistics about the movement and location of individuals, and in the frequency and locations of different interactions and activities (feeding, sleeping, grooming, fighting, etc.) They are willing to help in labeling relevant activities, even to the point of answering a few hundred quick questions per day of data (what's she doing here?), but they don't want to sit through 24 hours of video to do it. Ultimately they want an automated database that they can use to find out how many hours a day chimp Jane sleeps and where, histogram preferred eating locations, obtain statistics on who grooms whom, etc. Their equipment budget is an exceedingly generous $50K. P.S. If you want to "bar-code" the chimps, you have to do it in a way that will bother neither the chimps, nor the visitors to the zoo.
   - What kind of data do you want to extract?
   - How to collect such data?
   - How big data analytics can uncover the unexpected in your data?
   - How does the learning improve your system?
   - Draw the workflow of the proposed system and explain the process of the system.

PS-1

① odd man out:

let's have the array of unsorted integers be a[0], a[1], ... a[n-1] be the n integers.

→ Initialize K=1;

```
for ( i=0 ; i<n ; i++ )
{
    if (K==1)
    {
        for ( j=0 ; j<n ; j++ )
        {
            K=0 ;
            if ( i!=j)
            {
                if ( a[i] == a[j])
                {
                    K=1 ;
                    break
                }
            }
        }
    }
    else
    {
        K= i-1 ;
        break ;
    }
}
```

o/p : a[K], integer that appears only once.

(2) System Design: Big Data Applications.

Study of 6 chimpanzees - chimp island

(1) what kind of data do you want to extract?

- Video & audio data, sleep monitor data

(2) How to collect such data?

- Video & audio recording devices, sleep monitors

(3) How big data analytics can uncover the unexpected in your data?

- Analyaze the video data to findout the movement of chimpanzees using bar-code reader.
- measure the sleep data using sleep monitor data
- feeding, grooming & fighting can be identified by training the System on what does that mean.

(4) How does learning improve your System?

- Learning the System about each of the actions that chimpanzees perform will help in identifying those actions.
- Learning helps to get the answers quickly & makes the analysis easier.

(5) Draw the workflow of the proposed system and explain the process of the system?

(1) The monitoring devices like, audio & video recorder, sleep analyzer will record the data & store that information to a server - DB

(2) The system should be trained to identify the following.

- movement of chimpanzees & identify one another using bar-code
- Grooming & fighting activities based on video data

(3) The collected data is fed to the trained system and the data should be analyzed, segregated and the analysis will be stored in DB in a timely manner.

(4) Then the system is capable enought to reply to the question like. "what's she doing there?"