

Details of team members:

1. Shubham Jain, id:jainshubham0403@gmail.com
2. Vibhu Goel, id:goelvibhu1998@gmail.com
3. Kanishk Chandrayan, id:kk270398@gmail.com
4. Jaspreet Singh, id:jsk1961998@gmail.com

[Link for the single zip file containing all code solutions,readme and videos](#)

CELESTINI PROJECT INDIA 2018
TAKE-HOME EXAM
March 19, 2018

- 1. This exam has 5 questions. Write the answers in the space provided in the questions. Return your solutions in PDF format by 11:59PM (IST) on Mar 31, 2018 to email <celestiniprizeindia@gmail.com>.**
- 2. Submit all the code solutions in a single zip file or using a GitHub link. Provide a readme file to the code solution for each question.**

1. Multiple-choice questions (10 points)

Select one or more correct solutions. Please write your answer next to **Solution:**

A.) What types of learning, if any, best describe the following three scenarios:

- (i) A coin classification system is created for a vending machine. In order to do this, the developers obtain exact coin specifications from the U.S. Mint and derive a statistical model of the size, weight, and denomination, which the vending machine then uses to classify its coins.
- (ii) Instead of calling the U.S. Mint to obtain coin information, an algorithm is presented with a large set of labeled coins. The algorithm uses this data to infer decision boundaries which the vending machine then uses to classify its coins.
- (iii) A computer develops a strategy for playing Tic-Tac-Toe by playing repeatedly and adjusting its strategy by penalizing moves that eventually lead to losing.

- [a] (i) Supervised Learning, (ii) Unsupervised Learning, (iii) Reinforcement Learning
- [b] (i) Supervised Learning, (ii) Not learning, (iii) Unsupervised Learning
- [c] (i) Not learning, (ii) Reinforcement Learning, (iii) Supervised Learning
- [d] (i) Not learning, (ii) Supervised Learning, (iii) Reinforcement Learning
- [e] (i) Supervised Learning, (ii) Reinforcement Learning, (iii) Unsupervised Learning

Solution: [a]

B.) For an imbalanced dataset, which of the following metric/tool is not that useful?

- [a] F1 measure
- [b] Accuracy
- [c] Confusion Matrix
- [d] Precision

Solution:[b]Accuracy

C.) Consider the following implementation of a function `mysteryFunction` (pseudocode), where `x` is a positive integer:

```
mysteryFunction (x)
  xs = str(x)
  if len(xs) == 1
    return int(xs)
  n = int(xs[0]) + int(xs[1])
  if len(xs) == 2
    return n
  else
    return n + mysteryFunction(xs[2:])
```

What does `mysteryFunction(3223)` return

- [a] 0
- [b] 10

[c] 5

[d] 1

Solution:[b]10

D.) What is the output of the following program (in C) for input "Celestini Project"

```
#include "stdio.h"
int main()
{
    char arr[100];
    printf("%d", scanf("%s", arr));
    return 2;
}
```

[a] 0

[b] -1

[c] 1

[d] 2

Solution:[b]-1

E.) Which of the following options suggest the best approach to fix the high bias and high variance in a machine learning model? (Assume model has been trained on at least 1000 samples)

[a] To fix high bias, we can add more training samples; to fix high variance, we can reduce the number of training examples so it fits on them less

[b] To fix high bias, we can reduce our model's complexity; to fix high variance, we can increase our model's complexity

[c] To fix high bias, we can increase our model's complexity; to fix high variance, we can try reducing the number of features in the dataset

[d] To fix high bias, we can decrease the number of training samples; to fix high variance, we can increase the number of features in the dataset

Solution:[a]

F.) The major advantage(s) of prototyping over a Raspberry Pi over prototyping on a personal computer are

[a] cost

[b] faster processing speed

[c] small form factor

[d] low power consumption

Solution: [a],[c],[d]

G.) Which of the following statement(s) are correct?

- [a] A machine learning model with higher accuracy will always indicate a better classifier.
[b] When we increase the complexity of a model, it will always decrease the test error.
[c] When we increase the complexity of a model, it will always decrease the train error.

Solution:[b]

H.) What is the output of the program (in C)?

```
#include <stdio.h>
int main()
{
    int celestini[6] = {6,5,4,3,2,1};
    int *ptr = (int*)&celestini+1;
    printf("%d %d", *(celestini+1), *(ptr-1));
    return 0;
}
```

[a] 5 1

[b] 4 3

[c] 6 4

[d] 5 3

Solution:(a)

I.) A poor binary classification model for detecting a **rare** cancer disease *a/ways* predicts positive for presence of the disease. What can we infer about the model's performance?

- [a] The model has high accuracy, maximum precision but low recall.
[b] The model has poor accuracy, poor precision but maximum recall.
[c] The model has poor accuracy, maximum precision and minimum recall.
[d] The model has maximum accuracy, maximum precision but minimum recall.

Solution:[a]

J.) Which of the following problems are best suited for a machine learning approach?

- (i) Classifying numbers into primes and non-primes.
(ii) Detecting potential fraud in credit card charges.
(iii) Determining the time it would take a falling object to hit the ground.
(iv) Determining the optimal cycle for traffic lights in a busy intersection.

[a] (ii) and (iv)

[b] (i) and (ii)

[c] (i), (ii), and (iii).

[d] (iii)

Solution:[a]

2. Programming (10 points)

Given two sparse matrices A and B, perform multiply and convolution operation of the matrices in their sparse form itself. The result should consist of two sparse matrices, one obtained by multiplying the two input matrices, and the other obtained by convolution of the two matrices.

Recall that a sparse matrix is a matrix in which most of the elements are zero. Assume both the matrices are of size $N \times N$. Assume the number of non-zero elements in A and B are m_1 and m_2 respectively. Note that other entries of matrices will be zero as matrices are sparse.

Note: You may use any data-structure to represent the sparse matrix. The solution approach should not use in-built libraries for the multiplication or convolution of matrices.

(i) Write code to solve the above problem in Python, Java or C++

JAVA Code:

```
import java.util.Scanner;

public class celestnique2 {

    int data[][];
    int row, col;
    static int r1[], c1[], v1[], r2[], c2[], v2[];
    int len;

    public celestnique2(int r, int c, int size) {
        row = r;
        col = c;
        len = 0;
        data = new int[size][3];
    }

    public void insert(int r, int c, int val) {
        if (r > row || c > col) {
            System.out.println("Wrong entry");
        } else {
            data[len][0] = r;
            data[len][1] = c;
            data[len][2] = val;
        }
    }
}
```

```

        len++;
    }
}

public celestnique2 transpose() {

    celestnique2 result = new celestnique2(col, row, this.len);
    result.len = len;
    int count[] = new int[col + 1];
    for (int i = 1; i <= col; i++)
        count[i] = 0;
    for (int i = 0; i < len; i++)
        count[data[i][1]]++;
    int[] index = new int[col + 1];
    index[1] = 0;
    for (int i = 2; i <= col; i++)
        index[i] = index[i - 1] + count[i - 1];

    for (int i = 0; i < len; i++) {
        int rpos = index[data[i][1]]++;
        result.data[rpos][0] = data[i][1];
        result.data[rpos][1] = data[i][0];
        result.data[rpos][2] = data[i][2];
    }
    return result;
}

public void multiply(celestnique2 b) {
    b = b.transpose();
    int apos, bpos;
    celestnique2 result = new celestnique2(row, b.row, b.len);

    for (apos = 0; apos < len;) {

        int r = data[apos][0];

        for (bpos = 0; bpos < b.len;) {

            int c = b.data[bpos][0];

            int tempa = apos;
            int tempb = bpos;

            int sum = 0;
            while (tempa < len && data[tempa][0] == r && tempb

```

```

< b.len && b.data[tempb][0] == c) {

    if (data[tempa][1] < b.data[tempb][1])

        tempa++;

    else if (data[tempa][1] > b.data[tempb][1])

        tempb++;
    else

        sum += data[tempa++][2] *
b.data[tempb++][2];
    }

    if (sum != 0)
        result.insert(r, c, sum);

    while (bpos < b.len && b.data[bpos][0] == c)

        bpos++;
    }

    while (apos < len && data[apos][0] == r)

        apos++;
    }

    result.print();
}

public void convolve(int m1,int m2,int[] r12,int[] r22,int[] c12,int[]
c22,int[] v12,int[] v22,int n)
{
    int conv[]=new int[n];
    for(int i=0;i<m1;i++)

        for(int j=0;j<m2;j++)
        {
            // only if i+(m-i) equals m then convolution occurs
            //only r c elements needs to be considered in
sparse matrix else value is zero
            //dimensions of resultant convolution matrix is of
matrix A

            if(((r12[i]+r22[j])==n)&&(c12[i]+c22[j]==n))
            {

```



```

        conv[i]=v12[i]*v22[j];
    }

    }
    System.out.println("Convolution: " );
    System.out.println("Sparse Matrix: \nRow Column Value");

    for (int i = 0; i < m1; i++) {

        System.out.println(r12[i] + " "
                           + c12[i] + " " + conv[i]);
    }

}

public void print() {
    System.out.println("Dimension: " + row + "x" + col);
    System.out.println("Sparse Matrix: \nRow Column Value");

    for (int i = 0; i < len; i++) {

        System.out.println(data[i][0] + " " + data[i][1] + " " +
data[i][2]);
    }
}

public static void main(String args[]) {

    System.out.println("\nEnter n: ");
    Scanner reader = new Scanner(System.in);
    int n = reader.nextInt();
    System.out.println("\nEnter m1(no. of non zero elements in
matrix A: ");
    int m1 = reader.nextInt();
    System.out.println("\nEnter m2(no. of non zero elements in
matrix B: ");
    int m2 = reader.nextInt();
    celestnique2 a = new celestnique2(n, n,m1);
    celestnique2 b = new celestnique2(n, n,m2);

    r1 = new int[m1];
    c1 = new int[m1];

```

```

        v1 = new int[m1];
        r2 = new int[m2];
        c2 = new int[m2];
        v2 = new int[m2];

        for (int i = 0; i < m1; i++) {
            System.out.println("\nEnter Row Column Value of matrix
A:");

            r1[i] = reader.nextInt();
            c1[i] = reader.nextInt();
            v1[i] = reader.nextInt();
        }

        for (int i = 0; i < m2; i++) {
            System.out.println("\nEnter Row Column Value of matrix
B:");

            r2[i] = reader.nextInt();
            c2[i] = reader.nextInt();
            v2[i] = reader.nextInt();
        }

        for (int i = 0; i < m1; i++)
            a.insert(r1[i], c1[i], v1[i]);

        for (int i = 0; i < m2; i++)
            b.insert(r2[i], c2[i], v2[i]);

        reader.close();
        System.out.println("\nMultiplication: ");
        a.multiply(b);
            a.convolve(m1,m2,r1,r2,c1,c2,v1,v2,n);
    }
}

```

[Solution_Q2\(i\)](#)

(ii) What is the best time complexity of your solution (in terms of m_1, m_2, N)?

$O(N \cdot m_2 + N \cdot m_1)$

(iii) What is the best space complexity of your solution (in terms of m_1, m_2, N)?

$O(m_1 + m_2)$

3. Programming II (10 points)

Write an efficient algorithm that searches for a value in an $m \times n$ matrix. This matrix has the following properties:

- Integers in each row are sorted in ascending from left to right.
- Integers in each column are sorted in ascending from top to bottom.

For example,

Consider the following matrix:

```
[
  [1, 4, 7, 11, 15],
  [2, 5, 8, 12, 19],
  [3, 6, 9, 16, 22],
  [10, 13, 14, 17, 24],
  [18, 21, 23, 26, 30]
]
```

Given target = 5, return true.

Given target = 20, return false.

(i) Write code to solve the above problem in Python, Java or C++.

JAVA Code:

```
import java.util.Scanner;
```

```
public class celestniques3 {
```

```
    private static boolean search(int[][] mat, int m, int n, int x) {
```

```

        int i = 0, j = n - 1;

        while (i < m && j >= 0) {
            if (mat[i][j] == x) {
                return true;
            }
            if (mat[i][j] > x)
                j--;
            else
                i++;
        }
        return false;
    }

    public static void main(String[] args) {
        int mat[][] = { { 1, 4, 7, 11, 15 }, { 2, 5, 8, 12, 19 }, { 3, 6, 9, 16, 22 }, { 10, 13, 14,
17, 24 } };

        Scanner s = new Scanner(System.in);
        System.out.println("Enter the number to be searched:");
        int num = s.nextInt();
        boolean found = search(mat, 4, 5, num);
        if (found)
            System.out.print("Found");
        else
            System.out.print("Not found");
        s.close();
    }
}

```

[Solution_Q3\(i\)](#)

(ii) What is the best time complexity of your solution (in terms of m, n)?

Best Case- $O(1)$

Avg. & Worst Case- $O(n+m)$

(iii) What is the best space complexity of your solution (in terms of m , n)?

$O(m*n)$

4. Problem Solving (20 points)

Please select either problem 4A or 4B and provide your solution in detail. You may solve both problems for extra credit though it is not required.

4A. Cryptosystem Identifier (select either 4A or 4B)

Cryptography is associated with the process of converting plain text into unintelligible text and vice versa. The goal of problem is to identify the cryptosystem used in encrypting a given cryptogram using Support Vector Machine (SVM) and Back propagation Neural Networks (BPNN). We consider that the cryptogram are derived using Simple substitution or Vigenere.

[a] Simple substitution (SS) ciphers work by replacing each plaintext character by another one character. To decode cipher text letters, one should use reverse substitution and change the letters back.

[b] Vigenere cipher is a kind of polyalphabetic substitution cipher. It is about replacing plaintext letters by other letters. Parties have to agree on a common shared keyword (which may also be a sentence), which is used during encryption algorithm.

Data generation approach: Create 50 cryptograms by Simple Substitution (Key size: 26) and 50 cryptograms by Vigenere cryptosystems (key size: 3). Each of the cryptograms should be of size 200 characters consisting of only upper case alphabets and white spaces (i.e. total 27 characters).

You can use the following links for encoding

- Vigenere:

<https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/29443/versions/1/previews/VigenereDetails.html>

- Simple substitution:
<https://in.mathworks.com/matlabcentral/fileexchange/31522-substitution-cipher-encoder-and-decoder>

We are providing you with a dataset of ten plaintext, ten cryptograms by Vigenere, and ten cryptograms by simple substitution for testing your solution in the attachment (dataset_cryptosystem.doc)

Hint: You may consider using frequency pattern of the cryptograms for training the dataset.

(i) Write the solution for implementing Cryptosystem Identifier in MATLAB or Python. Give a brief description of what feature vectors you have used, how you designed the machine-learning model for SVM and BPNN, and what loss function did you use in each case.

(ii) Compare the performance of the classifiers based on SVM and BPNN using test samples. Did you use a validation approach on the dataset? What performance metric did you use to compare the performance? Why is this a good metric?

(iii) Plot the performance of your system for SVM and BPNN by varying parameters in your model.

You will be graded based on what you have submitted as well as your ability to explain your code.

4B. Designing IoT system (select either 4A or 4B)

Many applications such as robot navigation (wheeled robot for instance) require an estimate of where the obstacle is relative to the robot.

(i) Design a SONAR system using Arduino UNO that records the distance of the obstacle and the angle by which the sensor has rotated on the console.

Things you will require:

- Arduino UNO kit
(<https://www.amazon.in/Arduino-ATmega328P-ATMEGA16U2-Compatible-Cable/dp/B06XB81X82>)
- jumpwires
- breadboard/PCB boards
- ultrasonic sensor HC-SR04
(<https://www.amazon.in/Adraxx-HC-SR04-Ultrasonic-Distance-Measuring/dp/B01LXFUAEV>)

Answer:

Click here-->

[Link containing the code and demo video \(.rar file\) for Q4 B\) i\)](#)

(ii) Discuss the system you have designed with the following specifications:

[a] Explain the working principle behind the transceiver and how it measures the distance and angle

Answer:

1. This particular sensor transmits an ultrasonic sound that has a frequency of about 40 kHz. The sensor has two main parts- transducer that creates an ultrasonic sound wave while the other part listens to its echo.
2. The transmitter transmits the ultrasonic beam that strikes the flat surface of the obstacle and reflects back to meet the receiver. So, the sensor measures the time in which the transceiver emits the ultrasonic beam and receives it.
3. The distance is measured by multiplying the time duration with the speed of sound (340 metres/second) to give the distance in centimeters.

4. The angle by which the sensor gets rotated is measured by calculating the distance1 (BASE) and distance2 (Hypotenuse) and then applying cos inverse formula.

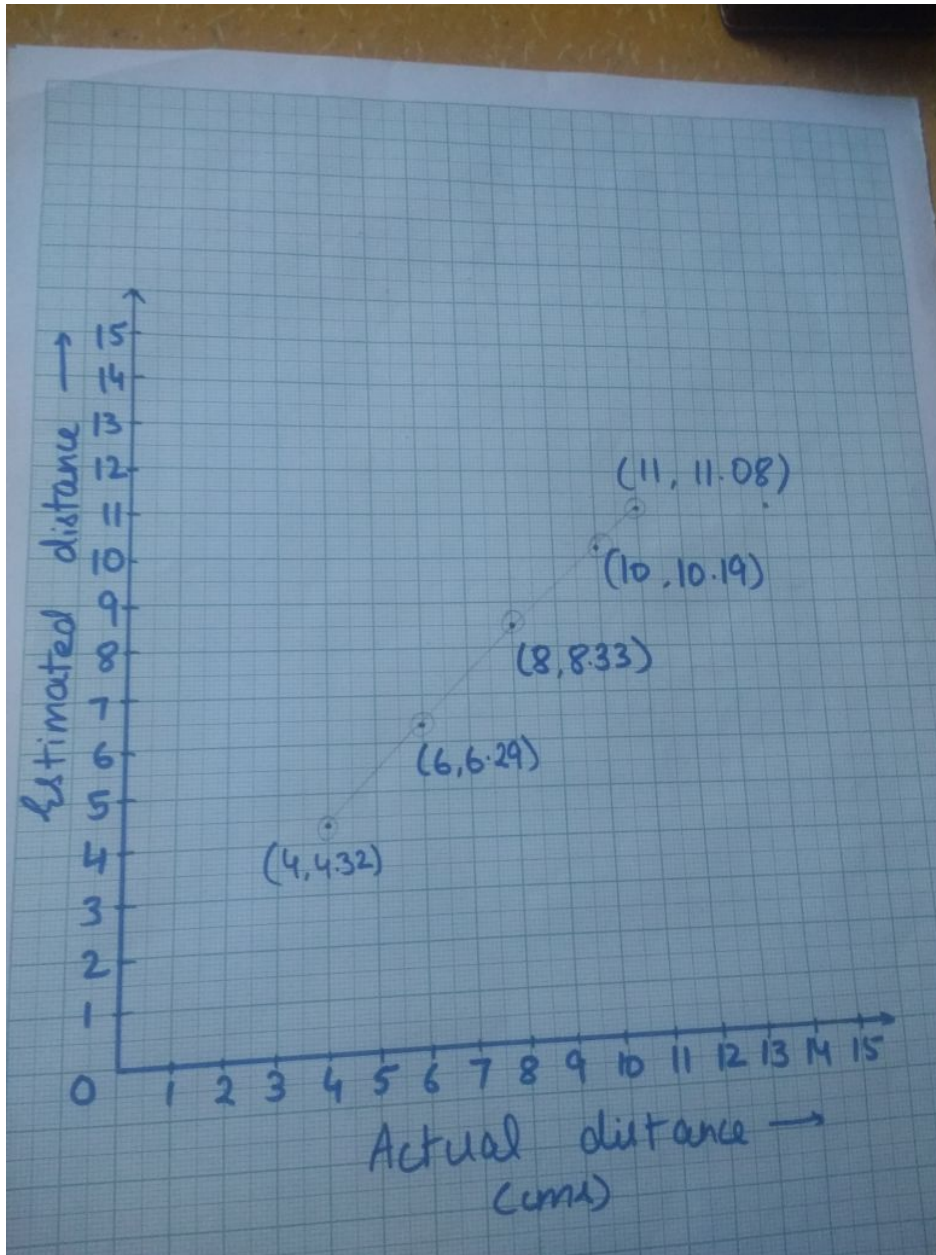
[b] Plot a graph between the estimated distance (y-axis) and actual distance (x-axis)

Answer:

TABLE:

Actual distance (X axis) (in cms)	Estimated distance (Y axis) (in cms)
4	4.32
6	6.29
8	8.33
10	10.19
11	11.08

GRAPH:



[c] Discuss any parameter which affects the performance of the system in the plot obtained in [b]

Answer:

1. The ultrasonic sensor gives its best output for the flat surfaces. But if the surface is curved or of any irregular shape, the sensor might not give the correct values because the ultrasonic beam is not reflected back properly hence affecting its functionality.
2. The obstacle needs to be in the rectilinear position with respect to the ultrasonic sensor in order to get detected, hence, if the obstacle lies beside/at side of the sensor, it might not get detected.

[d] Find the workable ranges of obstacle resolution (minimum and maximum size of the objects which can be detected)

Answer:

The minimum height of the obstacle should be at least 1.4 cms for the proper working of the ultrasonic sensor. The minimum width of the obstacle should be at least 0.7 cms.

There is no constraint on the maximum height and width of the obstacle. We can choose any obstacle no matter how bigger in size it is (though it should be bigger than the minimum dimensions required).

Submit this along with code files and readme in a .zip format or Github link. Also provide a demo video showing the results clearly.

(iii) Optional Part: Additional credits for novelty in circuit design (customised circuitry). Provide a blueprint of the circuit diagram using easyEDA (<https://easyeda.com/>) in case of customized circuitry. Can you construct a touch detection system using the same system which would convert it to give back the {x,y} coordinates of the point where touch is performed knowing the distance of the obstacle (finger in this case) and angle at which the sensor rotates. In case you give this a try include all necessary documentation and code files in .zip format.

Answer to the optional part:

[click here to download the .zip file having code and video for Q4B optional part](#)

5. Solving socio-economic problems using technology (10 points)

Select one of the two problems below:

- (i) Analytics and alerts on road safety using car mounted dashboard cameras
- (ii) Analytics and alerts on air pollution in Delhi using vision and IoT sensors

Discuss in about 500-600 words how you would design a solution for the problem you selected above. Your solution approach needs to consider the following parts:

- a) datasets or data acquisition for training
- b) choice of machine learning algorithm to run online or offline
- c) what platform can be used to run machine learning algorithm (for e.g. Raspberry Pi, smartphone, cloud)
- d) sending alerts over the network via peer-to-peer methods or cloud architecture.

This question is open-ended so you need to outline the design choices you will make. Include an architecture diagram and how you would measure the performance of the system you design. What demo can you show and what key challenges do you expect. (Note: Additional credits on out-of the box feasible and interesting ideas)

Answer:

Q5 i)

PROBLEMS/CHALLENGES FACED AND USE CASE:

Road safety is a major issue nowadays. Several steps have been taken by government but there are no such major outcomes of the same. As per the studies 1,46,377 people have died in 2017 in India due to road accidents out of which 9,317 is the contribution from foggy conditions .

VAAHAN is a device which focuses on reducing these road accidents to a major extent. It focuses on 4 major aspects...

TASK 1: During *FOGGY* and *SMOKY* conditions, there is a major problem of low vision which leads to collision of vehicles. People tend to look for resources or vehicles which they can use or follow to reach their destination. So, VAAHAN will be taking the responsibility to alert/notify the driver about the speed and distance of any vehicle in vicinity of the driver in order to prevent any sort of threat.

TASK 2 : In India, there are numerous cases when people lose their lives due to *DRINKING & DRIVING*. As per reports, an average of 7,061 people die due to drinking and driving per year in India. Hence, to prevent such incidents, VAAHAN will be involving an alert system associated with it that will be alerting the alcoholic driver to stop his/her vehicle if he/she is driving the car. Still, if the driver doesn't bother stopping the vehicle, VAAHAN will inform the police about the details of the driver and vehicle preventing any sort of threat to the driver or the people in vicinity of the vehicle.

TASK 3 : There are many cases when a person meets with an accident and there is no one around him/her to provide AID at the point of incident. But from now on, VAAHAN will provide a channel to deal with such incidents. It will be sending the vehicle's current location to the near and dear ones of the driver whenever any mishap occurs to alert them about the incident INSTANTANEOUSLY so that some AID could be provided to the person in need of help.

TASK 4: VAAHAN will also be targeting to alert the driver if any vehicle approaches it/or if the driver is approaching any vehicle, so that driver can take required action like making adjustments in the speed of the vehicle to prevent accident. It will also be collecting the evidence of any sort of accident that happened by recording a video of the same.

PLATFORM USED AND IMPLEMENTATION:

We are using **IoT** and **machine learning** as major platforms to develop VAAHAN.

VAAHAN will be using RASPBERRY PI 3 as its major component that will be responsible for handling all the tasks of our device.

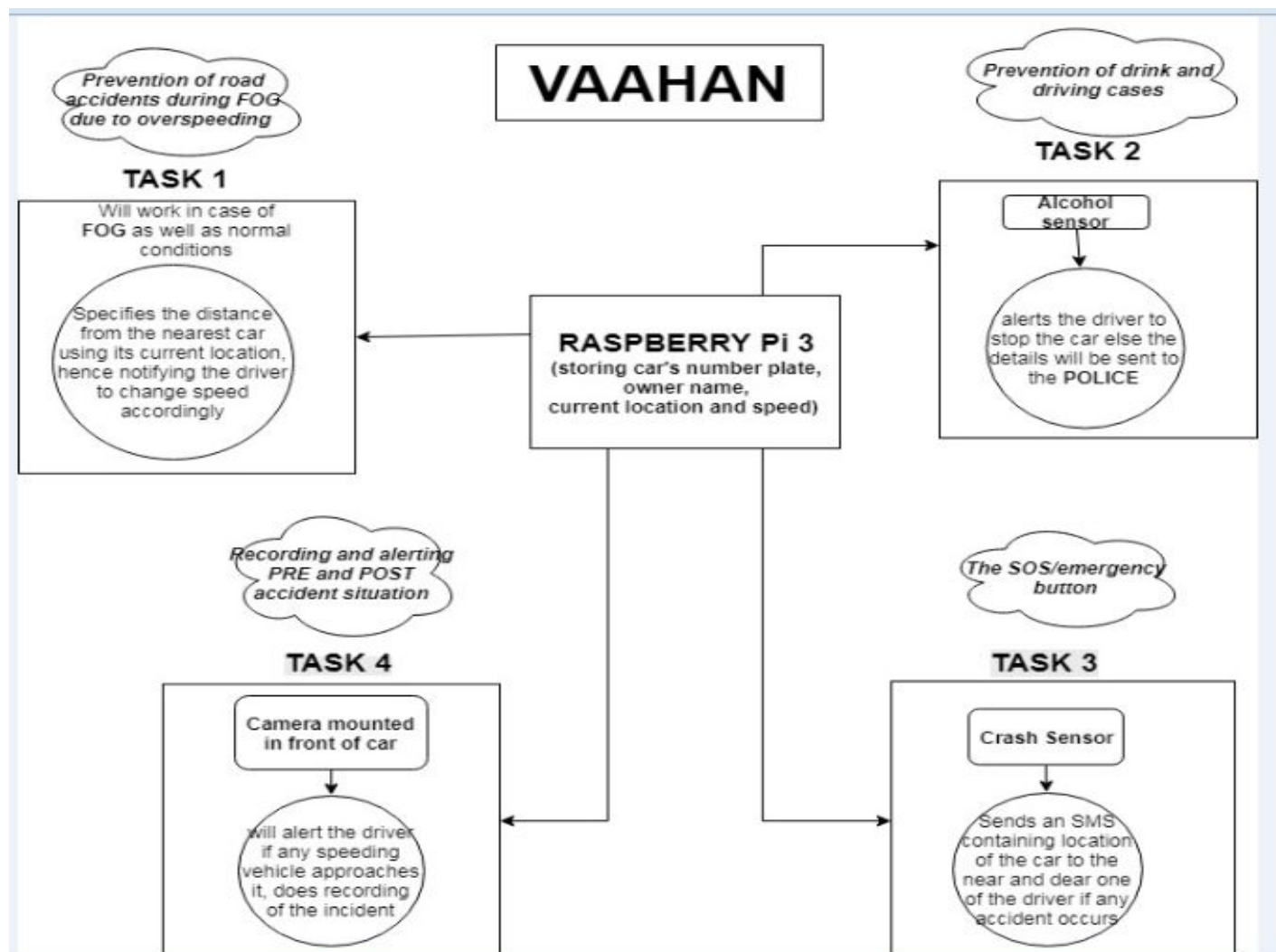
It will be integrated with a GPS module due to which it will be able to notify the driver about other vehicles (which have installed our system in their vehicles) in his/her vicinity even during foggy conditions.

It will also be integrated with an alcohol sensor and GSM module that will sense and alert the police if the driver is drinking while driving.

The same GSM and GPS module will be used to inform the near and dear ones of the driver if any mishap has occurred. The SOS button will be triggered automatically whenever the crash sensor (already embedded inside the vehicle) gets activated.

It will also have a camera as its part that will be used to notify/alert the driver about any vehicle approaching it using machine learning . This task does not require active internet connection. The camera will be taking new videos (after every 20 seconds by replacing the previous one) that will help to collect the post accident evidences (like the number plate of the incoming vehicle,etc.) if needed.

ARCHITECTURE DIAGRAM:



LINK FOR DEMO VIDEO: [DEMO VIDEO:](#)