

# **ROAD ACCIDENT PREDICTION**

## **PHASE-1**

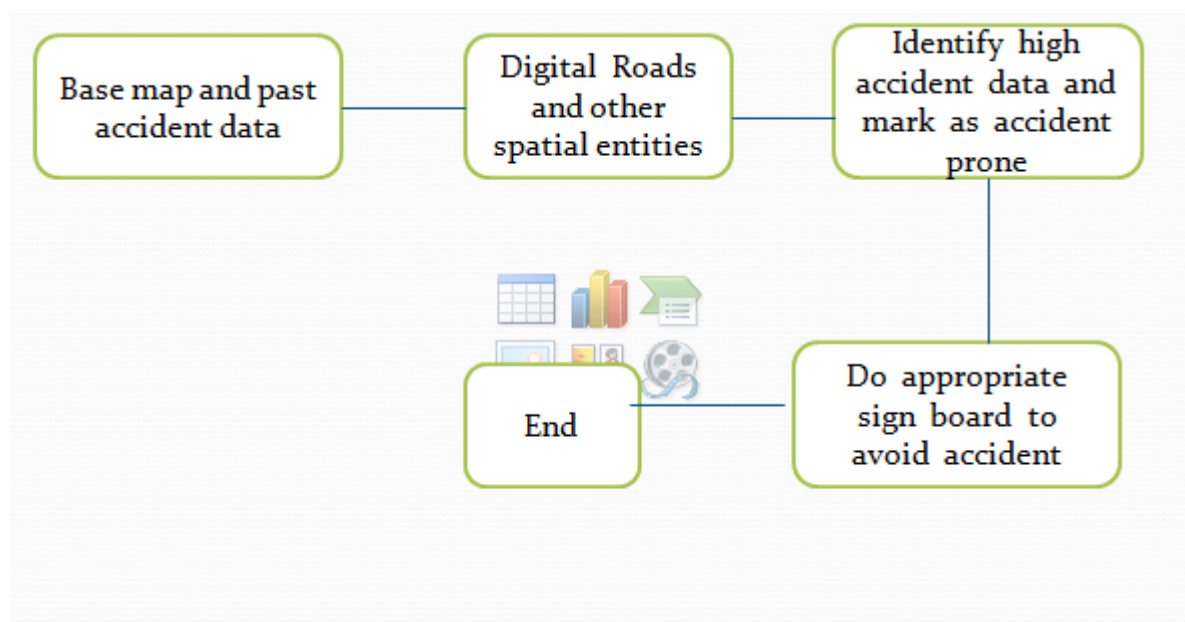
Submitted by :

S. Srisha – 2017103042

## **INTRODUCTION :**

This project is done using python language , We built a classifier to train on 80% and test on 20% of accident prediction of UK dataset. We used Random Forest Classifier and Decision tree Classifier to predict a output. We used Seaborn library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. The dataset holds 36314 rows . We enfolded a data for the Random Forest Classifier and predicted certain results to it.

## **BLOCK DIAGRAM :**



## **STEPS :**

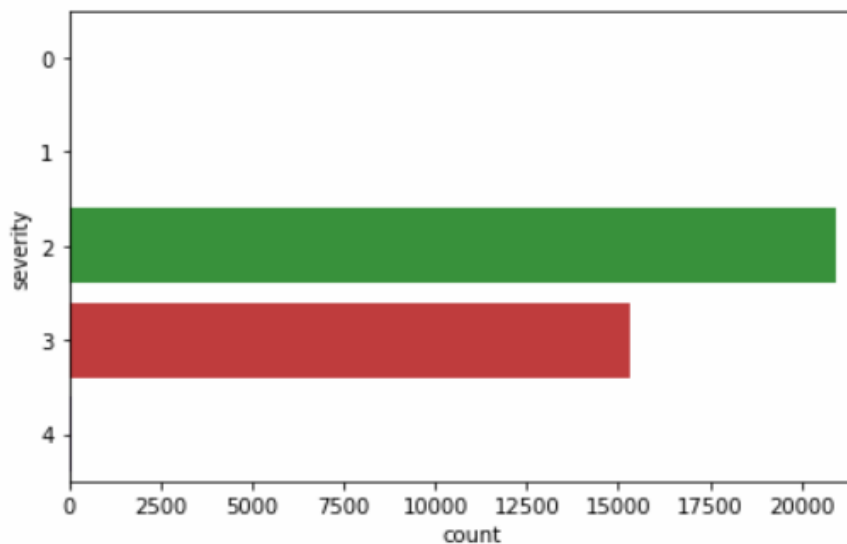
1. Importing the dataset file for the given constraints.

It prints the Top 5 rows of dataset File as i called the function called head() in my code.

	wkt_geom	code	severity	time	lat	desc	lon	name	clust	NN
0	POINT(6119950.59451242536306381 1886366.970110...	201	2	04/10/13 08:00 AM	77.165655	NaN	26.984052	Accident(s)	31	11
1	POINT(6163265.98209197819232941 2195198.763392...	201	2	04/10/13 05:30 PM	78.015568	NaN	27.117024	Accident(s)	14	3
2	POINT(6156011.13359668850898743 2064637.280291...	201	3	04/10/13 05:55 PM	77.656750	NaN	27.098630	Accident(s)	11	52
3	POINT(6040154.25808104187995195 2127971.038348...	201	3	04/10/13 06:17 PM	77.825210	NaN	26.694260	Accident(s)	28	11
4	POINT(6160755.65850416570901871 2058394.601888...	243	3	04/10/13 06:19 PM	77.639801	NaN	27.115337	Accident. Left lane blocked	11	51

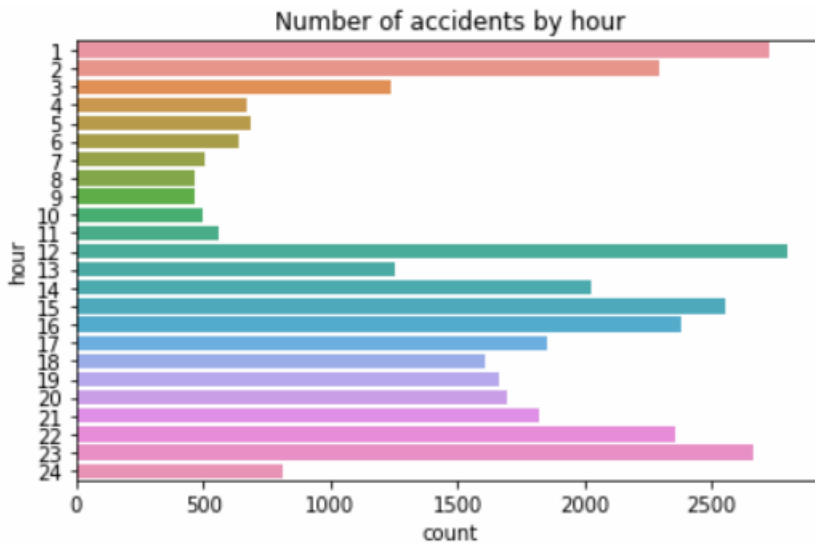
## 2.Characterstic graph:

It visualises the graph between Severity and Count. The Severity is represented in form of numbers ranging from 0 to 4.



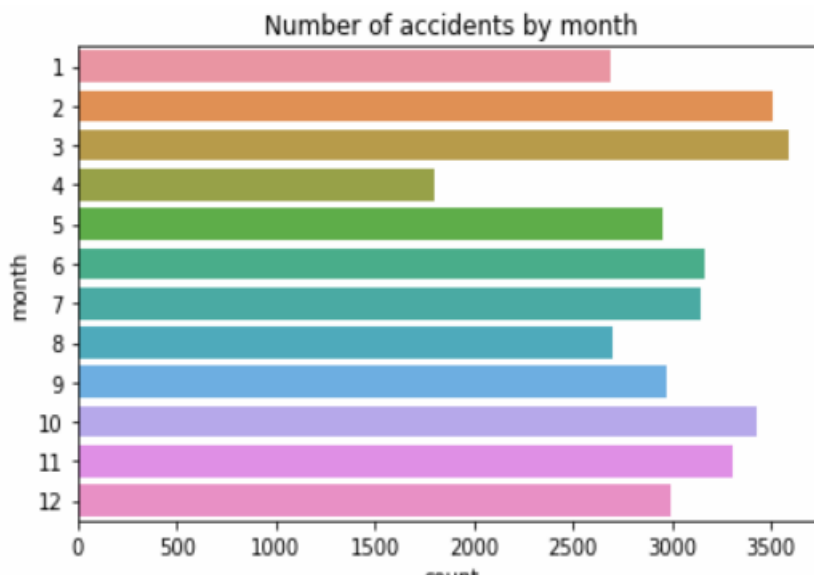
## 3. Classification graph :

It visualises the graph between Number of accidents by hour and count.



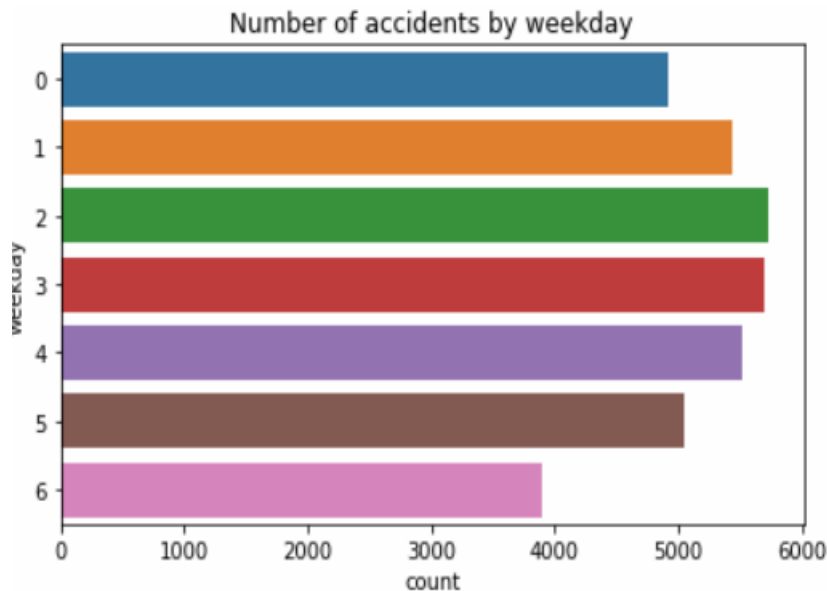
#### 4. Classification graph

This graph usually calculates the graph statistics here it takes number of accidents by month vs Count.



#### 5. Classification graph :

This graph usually calculates the graph statistics here it takes number of accidents by Weekday vs Count.



### Error's and Outlier removing :

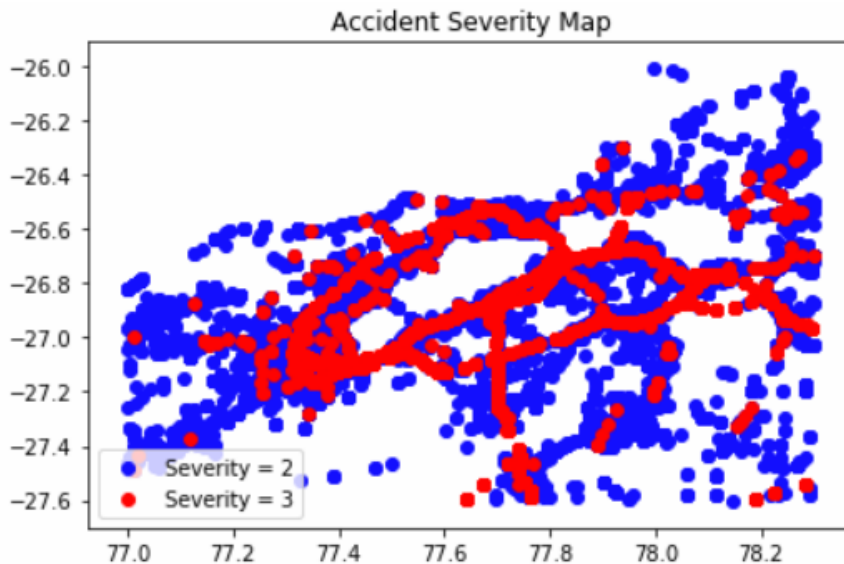
Here we check for outliers and boundaries. Outliers are extreme values that deviate from other observations on data, they may indicate a variability in a measurement, experimental errors or a novelty.

count	
2	20944
3	15296
4	54
1	19
0	1

### Distributions of features and labels :

Plotting of the dataset with a different color depending on the severity (2,3)

The mapping is between longitude and latitude by representing it with different colours i.e., Blue for severity = 2 and Red for severity = 3.



### Applying Cross Validation and Splitting the data into training set and test sets , ratio of 2:8

We assign test size as 20% of actual data set , random state is set as 42 . Also we defined random state to get consistent and same results , regardless of the training iterations , So that the values in the train and test sets are homogenous.

### Basic Algorithm Model

Worst Accuracy for my model :

worst accuracy: 0.577924944812362

### Applied algorithm's

## Random Forest Algorithm

Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object.

### N-FOLDING THE DATA FOR RANDOM FOREST CLASSIFIER

TOTAL=36000 ,

TRAINING DATA(80%)=28800 ,

TESTING DATA(20%)=7200 ,

SPLITTING DATA(10%)=3600

PARTITION No.	TRAINING DATA	TESTING DATA
1)	0-28799	28800-36000
2)	3601-32399	0-3600 , 32400-36000
3)	7201-36000	0-7200
4)	10801-36000 , 0-3600	3601-10800
5)	14401-36000 , 0-7200	7201-14400
6)	18001-36000 , 0-10800	10801-18000
7)	21601-36000 , 0-14400	14401-21600
8)	25201-36000 , 0-18000	18001-25200
9)	28801-36000 , 0- 21600	21601-28800
10)	32400-36000 , 0-25200	25201-32400

PARTITION 1 :

Random Forest Algorithm

Key Metrics :

```
----- RandomForestClassifier .
recall : 0.87
precision : 0.83
f1 : 0.85
accuracy : 0.87
-----
```

PARTITION 2 :

Random Forest Algorithm

Key Metrics :

```
----- RandomForestClassifier ----
recall : 0.92
precision : 0.86
f1 : 0.89
accuracy : 0.90
-----
```

---

PARTITION 3:

Random Forest Algorithm

Key Metrics :

```
----- RandomForestClassifier -----
recall : 0.90
precision : 0.86
f1 : 0.88
accuracy : 0.90
-----
```

PARTITION 4 :

Random Forest Algorithm

Key Metrics:



```
----- RandomForestClassifier ---
recall : 0.78
precision : 0.79
f1 : 0.79
accuracy : 0.82
-----
```

PARTITION 5 :

Random Forest Algorithm

Key Metrics:

```
----- RandomForestClassifier ---
recall : 0.82
precision : 0.81
f1 : 0.81
accuracy : 0.85
-----
```

---

PARTITION 6 :

Random Forest Algorithm

Key Metrics:

```
----- RandomForestClassifier ---
recall : 0.85
precision : 0.80
f1 : 0.83
accuracy : 0.86
-----
```

PARTITION 7 :

Random Forest Algorithm

Key Metrics:

---

```
----- RandomForestClassifier -----
recall : 0.77
precision : 0.84
f1 : 0.80
accuracy : 0.84
-----
```

---

PARTITION 8 :

### Random Forest Algorithm

#### Key Metrics :

```
----- RandomForestClassifier ----
recall : 0.86
precision : 0.84
f1 : 0.85
accuracy : 0.86
-----
```

---

PARTITION 9 :

### Random Forest Algorithm

#### Key Metrics :

```
----- RandomForestClassifier -----
recall : 0.87
precision : 0.81
f1 : 0.84
accuracy : 0.86
-----
```

PARTITION 10 :

### Random Forest Algorithm

#### Key Metrics :

```
#printScores(y_test, y_pred, RandomForestClassifier)
```

```
----- RandomForestClassifier -----  
recall : 0.86  
precision : 0.84  
f1 : 0.85  
accuracy : 0.87  
-----
```

---

## Mean for overall partitions :

Overall mean for keymetrics

Recall : 0.863

Precision : 0.828

Accuracy : 0.863

## Tree Algorithm:

DECISION TREE :

This is Structured tree where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves.

## N-FOLDING THE DATA FOR DECISION TREE ALGORITHM:-

PARTITION 1

### Key Metrics :

```
----- tree -----  
recall : 0.88  
precision : 0.88  
f1 : 0.88  
accuracy : 0.90  
-----
```

## PARTITION 2

### Key Metrics :

```
----- tree -----  
recall : 0.91  
precision : 0.89  
f1 : 0.90  
accuracy : 0.92  
-----
```

## PARTITION 3

### Key Metrics :

```
----- tree -----  
recall : 0.91  
precision : 0.89  
f1 : 0.90  
accuracy : 0.91  
-----
```

## PARTITION 4

### Key Metrics :

```
----- tree -----  
recall : 0.85  
precision : 0.83  
f1 : 0.84  
accuracy : 0.86  
-----
```

## PARTITION 5

### Key Metrics :

```
----- tree -----  
recall : 0.89  
precision : 0.84  
f1 : 0.87  
accuracy : 0.89  
-----
```

## PARTITION 6

### Key Metrics :

```
----- tree -----  
recall : 0.87  
precision : 0.88  
-----
```

```
f1 : 0.88
accuracy : 0.90
-----
```

## PARTITION 7

### Key Metrics :

```
----- tree -----
recall : 0.87
precision : 0.87
f1 : 0.87
accuracy : 0.89
-----
```

## PARTITION 8

### Key Metrics :

```
----- tree -----
recall : 0.90
precision : 0.87
f1 : 0.88
accuracy : 0.89
-----
```

## PARTITION 9

### Key Metrics :

```
----- tree -----
recall : 0.88
precision : 0.87
f1 : 0.88
accuracy : 0.89
-----
```

## PARTITION 10

### Key Metrics :

```
----- tree -----
recall : 0.88
precision : 0.87
f1 : 0.87
accuracy : 0.89
-----
```

## Mean for overall partitions :

Overall mean for keymetrics

Recall : 0.884

Precision : 0.869

Accuracy : 0.894