

# **Predicting Car accident severity**

**Anuj Litoriya**

**September 23**

## **1. Introduction:**

### **1.1 Background**

Street traffic wounds are presently assessed to be the eighth driving reason for death over all age bunches internationally, and are anticipated to turn into the seventh driving reason for death by 2030. Investigating a critical scope of elements, including weather conditions, extraordinary occasions, road works, gridlocks among others, an exact forecast of the severity of the accidents can be performed.

These experiences, could permit law authorization bodies to distribute their assets all the more successfully ahead of time of possible accidents, forestalling when and where an extreme accidents can happen just as sparing both, time and cash. What's more, this information on a serious mishap circumstance can be cautioned to drivers so they would drive all the more cautiously or even change their course on the off chance that it is conceivable or to medical clinic which could have set everything prepared for an extreme intercession ahead of time.

### **1.2 Problem**

Predicting the car accident t severity based on multiple factors such as weather condition, area, work roads , extraordinary occasions , gridlocks , traffic etc.

### **1.3 Interest**

Government ought to be exceptionally intrigued by precise predictions of the severity of a mishap, so as to decrease the hour of appearance and in this manner spare a lot of individuals every year. Others intrigued could be private organizations putting resources into advances meaning to improve street safeness.

## **2. Data Section**

### **2.1 Data source**

Data has been imported from following page of kaggle

<https://www.kaggle.com/ahmedlhlou/accidents-in-france-from-2005-to-2016>

### **2.2 Feature Selection**

The information comprising of the apparent multitude of recorded accidents in France from 2005 to 2016. The qualities informational index contains data on the time, spot, and sort of crash, weather and lighting conditions furthermore, sort of convergence where it happened. The spots informational index has the street specifics, for example, the inclination, shape and class of the street, the traffic system, surface conditions and foundation. On the client informational index it very well may be discovered the spot involved by the clients of the vehicle, data on the clients associated with the reason of voyaging, severity of the utilization of security gear furthermore, data on the people on foot. The vehicle informational collection contains the ow and kind of vehicle, and the occasion one names the accidents happening in a vacation. All informational collections share the identification number.

### **2.3 Description**

The dataset that came about because of the component choice comprised in 839,985 samples, every one depicting a mishap and 29 different highlights.

### **2.4 Data Cleaning**

The data cleaning is the way toward giving an appropriate organization to the data for its further analysis. The progression was to manage missing values and outliers. Initially the scope, longitude and street number were dropped structure the data 3 edge as in excess of a half of its values were NaN or 0 which is an exception in this case.

With respect to kind of the data, all highlights had an intelligible data type aside from for the date include which was de need with the string type. I utilized the to data capacity of pandas to de ne the date include with the date time type. All things considered, 24 highlights remained.

### 3. Exploratory Data Analysis

To begin with, the appropriation of the objective's values was visualized. The plot affirmed that it is a fair labeled dataset as the examples are partitioned 56-54 with more instances of lower severity. Then an irregularity analysis was performed, picturing the worldwide pattern of every day accidents just as the measure of accidents gathered by years, month of the year, and day of the week

```
Out[22]: <matplotlib.legend.Legend at 0x15200384df0>
```

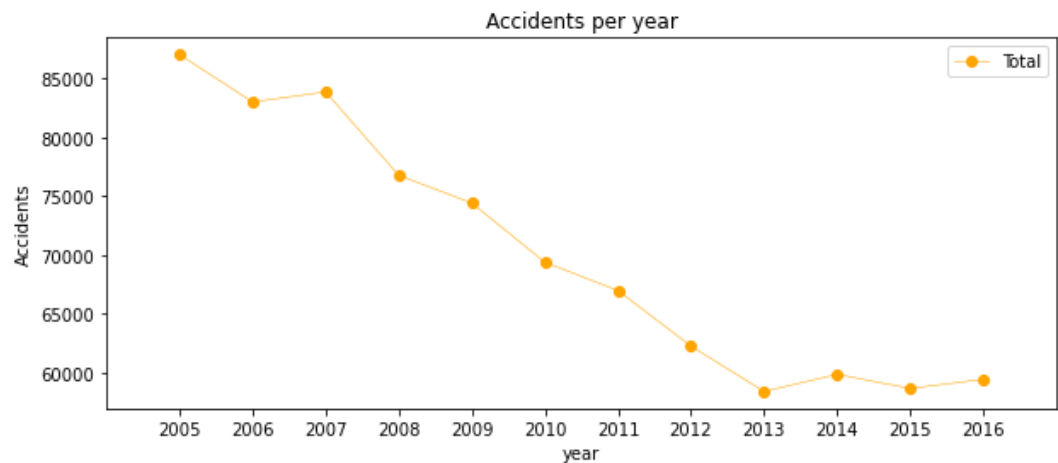


Fig. 1: Lineplot of total amount of accidents per year.

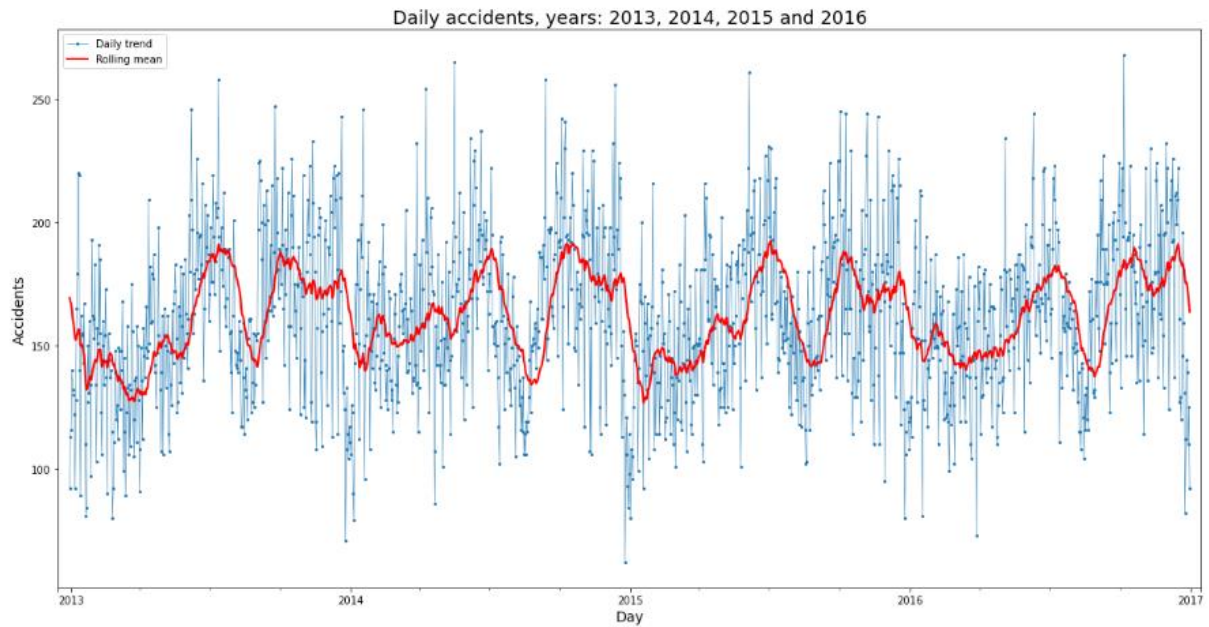


Fig 2: Line plot- accident per day during the 2013, 2014, 2015 and 2016. includes the rolling mean, window size 30 days

Out[20]: <matplotlib.legend.Legend at 0x1ed02e14430>

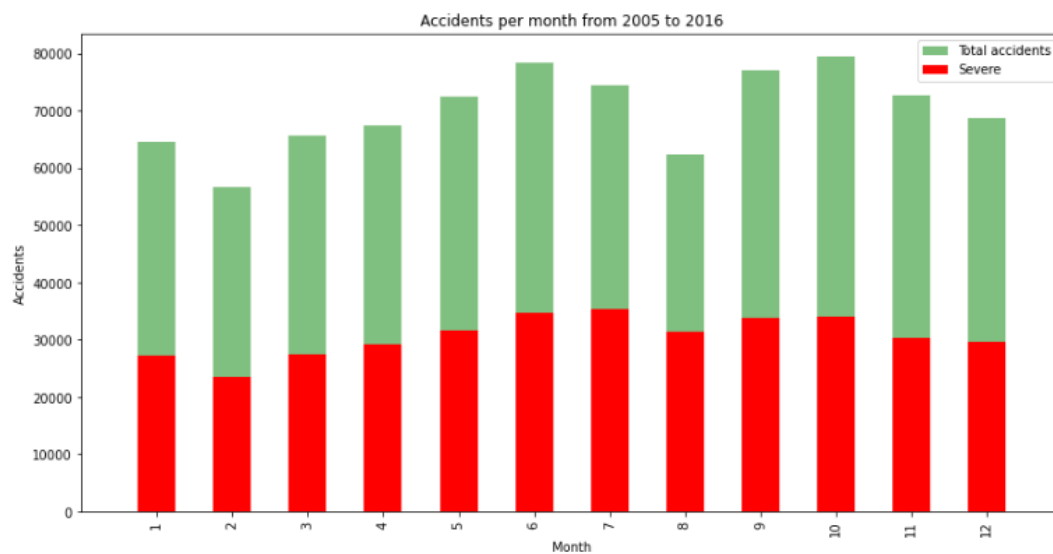


Fig 3: Bar plot - accident per month from 2005 to 2016.

Out[21]: <matplotlib.legend.Legend at 0x1ed030bd310>

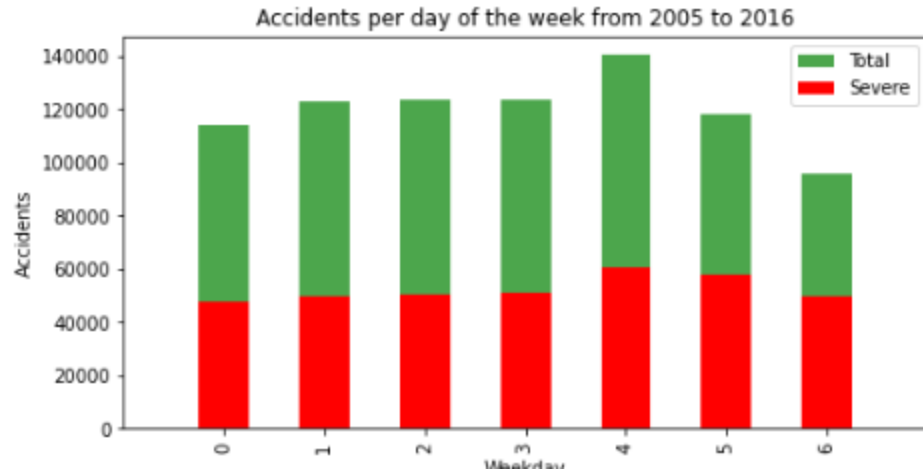


Fig 4: Bar plot - accident per day of the week from 2005 to 2016.

Out[22]: 839985

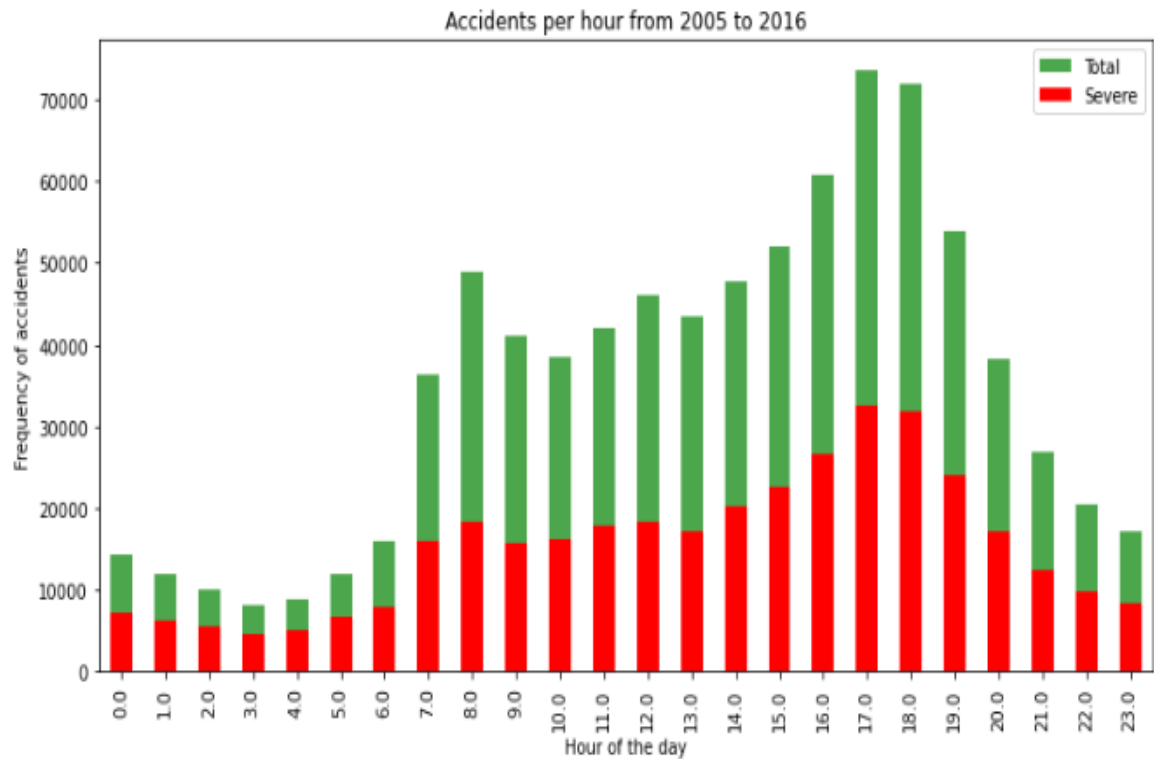


Fig 5: Line plot- of accidents per year

## 4. Modeling & Prediction

Multiple classification algorithms have been tuned and worked for the prediction of the degree of mishap severity. These algorithms gave a supervised learning approach foreseeing with certain accuracy and computational time. These two properties have been contrasted all together with decide the most appropriate calculation for his particular issue .Firstly, the 839.985 lines where part 80/20 between the training and test sets, subsequently an extra 80/20 split was performed among the training tests making the approval set for the improvement of the models. Then the data was normalized giving zero mean and unit difference to all highlights.

The decision tree model was moved up to the random forest. With the default random forest the highlights were arranged by pollution based significance in the prediction of the severity. In this manner, the 10 least significant highlights were dropped to diminish the calculation multifaceted nature for the KNN and SVM models. Keeping with 13 highlights the accuracy remained the equivalent and the computational time diminished essentially. In the wake of assessing the boundaries for every calculation these were the models

### Decision Tree:

```
> print( Accuracy : ,score_tree)
Time taken : 17.06349468231201
Accuracy : 0.637048170359678
```

### Random Forest

```
> print( Accuracy : ,score_rf)
Time taken : 254.37921977043152
Accuracy : 0.7182770576943109
```

#### 4.1. Confusion Matrix:

```
Time taken : 40.87483882904053
Jaccard : 0.7232391054602165
      precision    recall  f1-score   support

     0       0.73     0.81     0.77     94297
     1       0.72     0.61     0.66     73700

 accuracy          0.72     167997
 macro avg         0.72     0.71     0.71     167997
weighted avg         0.72     0.72     0.72     167997
```

### Random Forest

Time taken : 11.575872421264648

Jaccard : 0.6618332470222682

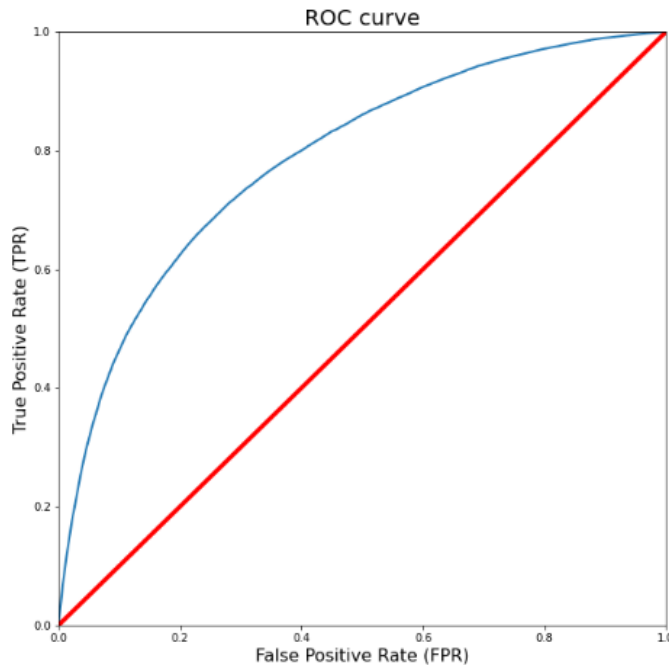
	precision	recall	f1-score	support
0	0.66	0.82	0.73	94297
1	0.67	0.46	0.54	73700
accuracy			0.66	167997
macro avg	0.66	0.64	0.64	167997
weighted avg	0.66	0.66	0.65	167997

Logistic Regression

## 5. Results

- According to the accuracy , precision and Recall , Random Forest is the best suited Model for this particular classification.
- It actually improved accuracy from 63% to 72%.





## 6. Observation

The accuracy that I am able to predict for this case is nearly 71% . There are multiple factors which can improve the accuracy like Speed , time of travelling etc can be significantly used for the further and efficient prediction . There are still lot of characteristics that are not predicted but I think as the evolution of tech is happening this will be done as an effective way.

The problem which I found was that the target labels in this classification are given as low or high . if that can be converted into a scale of 0- 100 than a regression model can be built with more efficient way.

Also for the next step or the update of this project can be done by adding the functionality of accident prediction by determining the spot and the area.

## 7. Conclusion

In this Particular case study , I worked on the data set of accident in France from 2005 - 2016. In which I analyzed the relationship between seriousness of an accident and a few qualities which portray the circumstance that included the accident .Initially I felt that highlights, for example, environmental conditions, the lighting or being an occasion would be the most pertinent ones, yet I recognized the office the day and season of the accident,

the street class and kind of crash among the most significant highlights that a demonstration to the gravity of the accident.