Predicting Traffic Accident Severity

Coursera Capstone Project

Ghenima CHELLOUCHE

Business Understanding

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

The aim of this project is to determine the possibility we get into a car accident and how severe it would be using these data

French road safety observatory and French government would be very interested to predict the severity of an accident

Data

Data Source :
Kaggle(https://www.kaggle.com/ahmedlahlou/accidents-in-france-from-2005-to-2016)

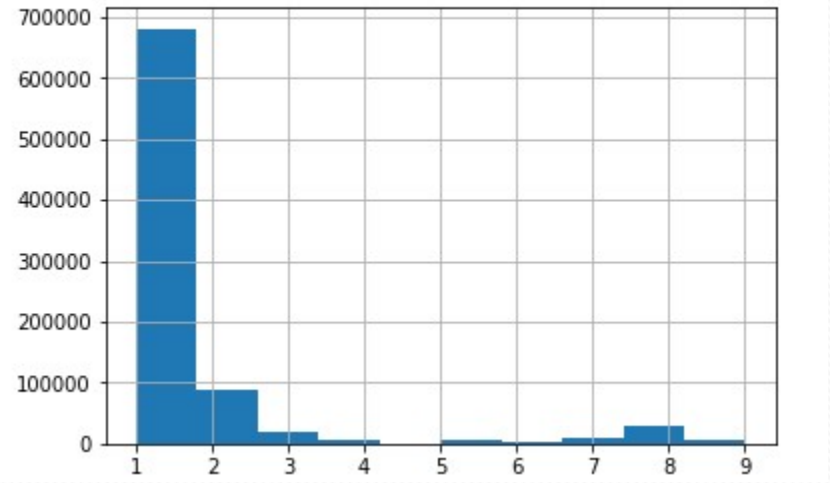
After Feature selection and data cleaning we have a data set with 839983 rows and 17 features.

Methodology(1.1)

The “atm” attribute represent the atmospheric condition and after analysis we can find that we have more accident when atmospheric condition is normal.

```
accidents["atm"].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x20db99c6908>
```

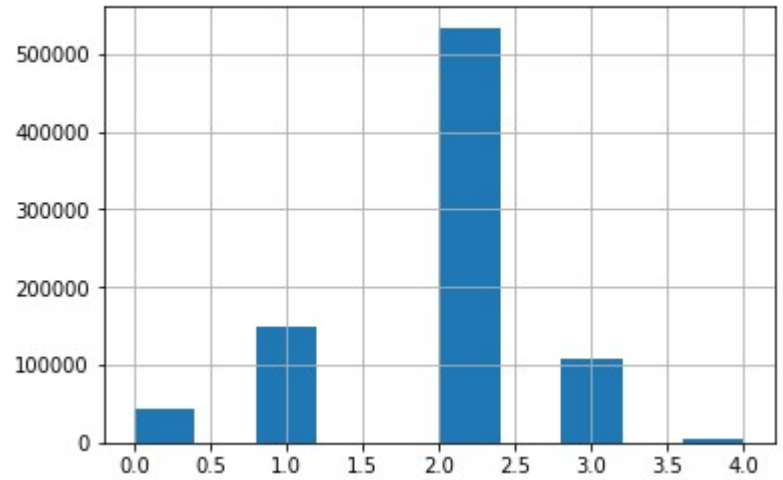


Methodology(1.2)

The “circ” attribute represent the traffic regime, after analysis we can find that we have more accident when the traffic regime is bidirectional.

```
accidents["circ"].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x20d8023cdd8>
```



Methodology(1.3)

Model development :

Random Forest : 100 decision trees
K-nearest neighbors : K=13
Logistic Regression : C = 0.01, solver = liblinear

Results

Algorithm	Accuracy
KNN	0.6796
Random Forest	0.7324
Logistic Regression	0.6626

The accuracy of the random Forest model equal to 0.732 is the better one comparing it to the accuracy of KNN equal to 0.6796 and Logistic Regression equal to 0.6626.

Conclusion

In this study, I tried to predict the severity of accident using different attributes. In first time I tried to analyze and understand the data set I choose to make my study. In second time I tried to makes models and to find the best model presenting a best accuracy.