

# Coursera Capstone Project

## Predicting traffic accidents severity

Presentation by : Tarakant Seth

“ 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. ”

Predicting the accident severity in advance could be used to send the exact required staff and equipment to the place of the accident, thus saving a significant amount of lives each year.

Road safety should be a prior interest for governments, local authorities and private companies investing in technologies that can help reduce accidents and improve overall driver safety.

# Data used

Kaggle: recorded accidents in France from 2005 to 2016

# EDA-Target

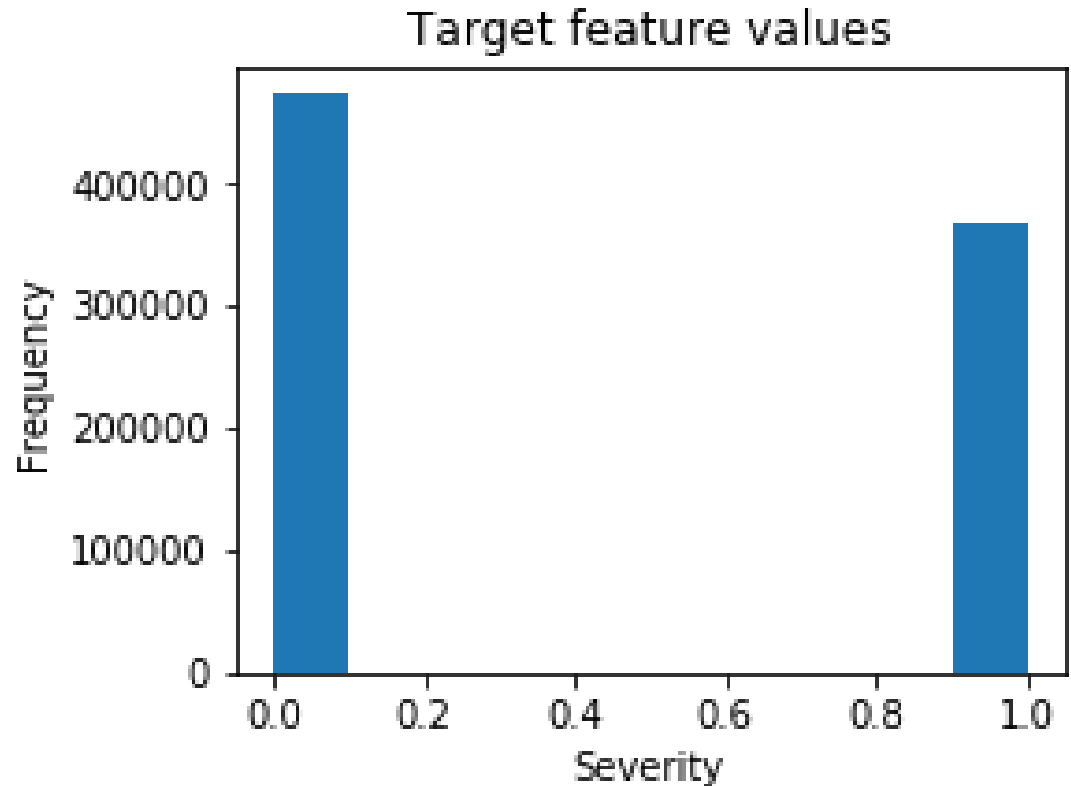
The target feature a binary classifier, describing the accident severity.

0: low severity.

1: high severity, from

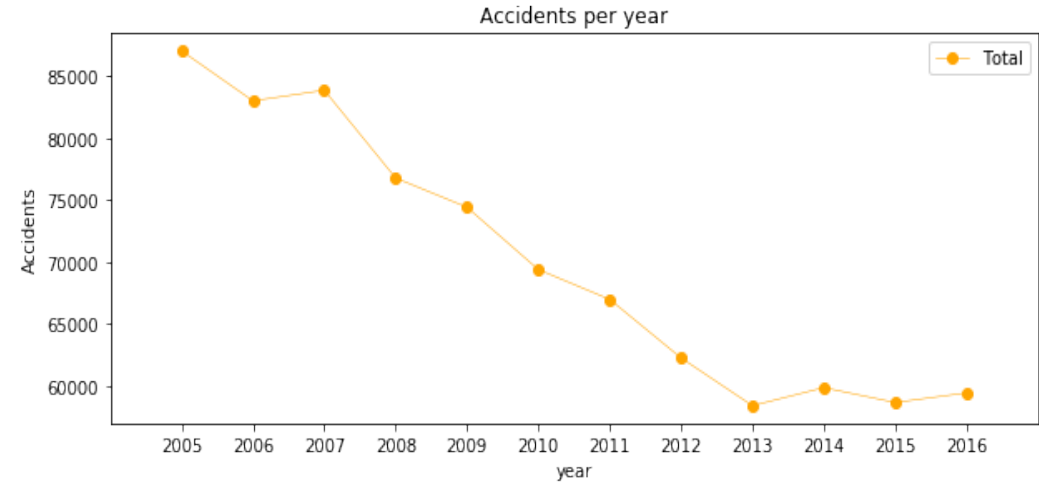
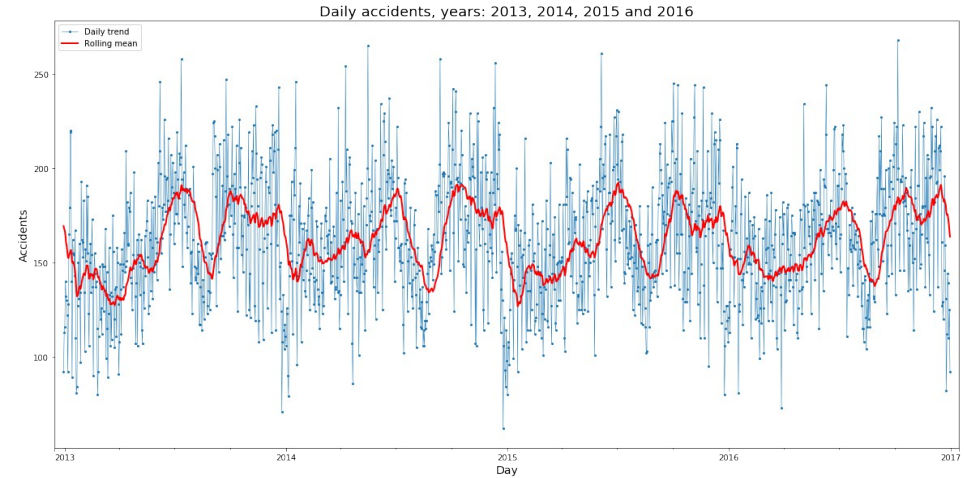
hospitalized wounded injuries to death.

It is a balanced labeled dataset with more cases of lower severity.



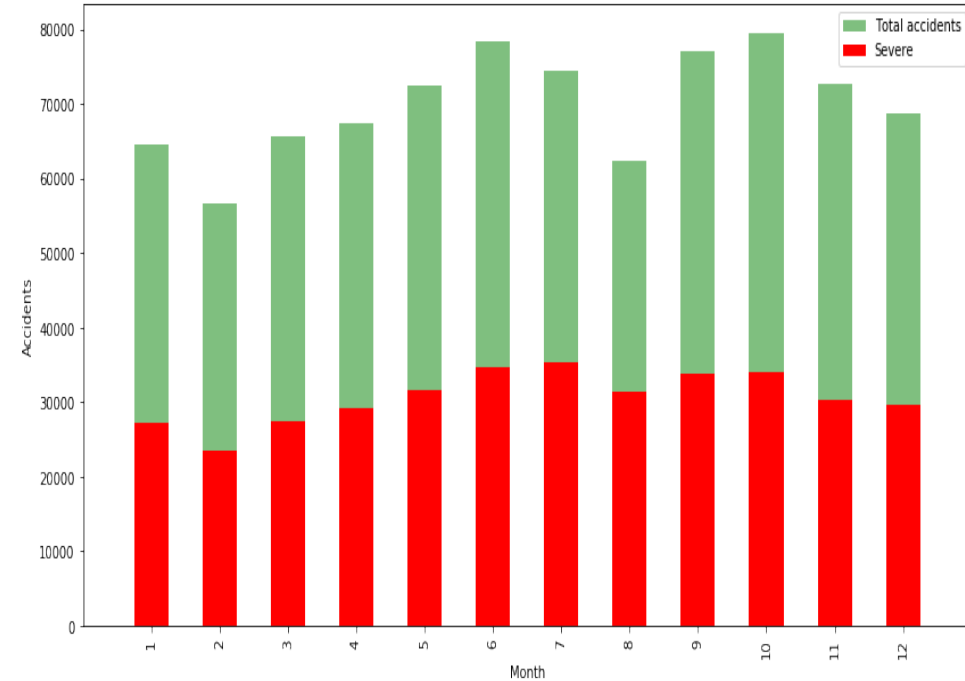
# EDA-Seasonality

The number of traffic accidents decreased over the years from 2005 to 2013, after which the trend became stable.

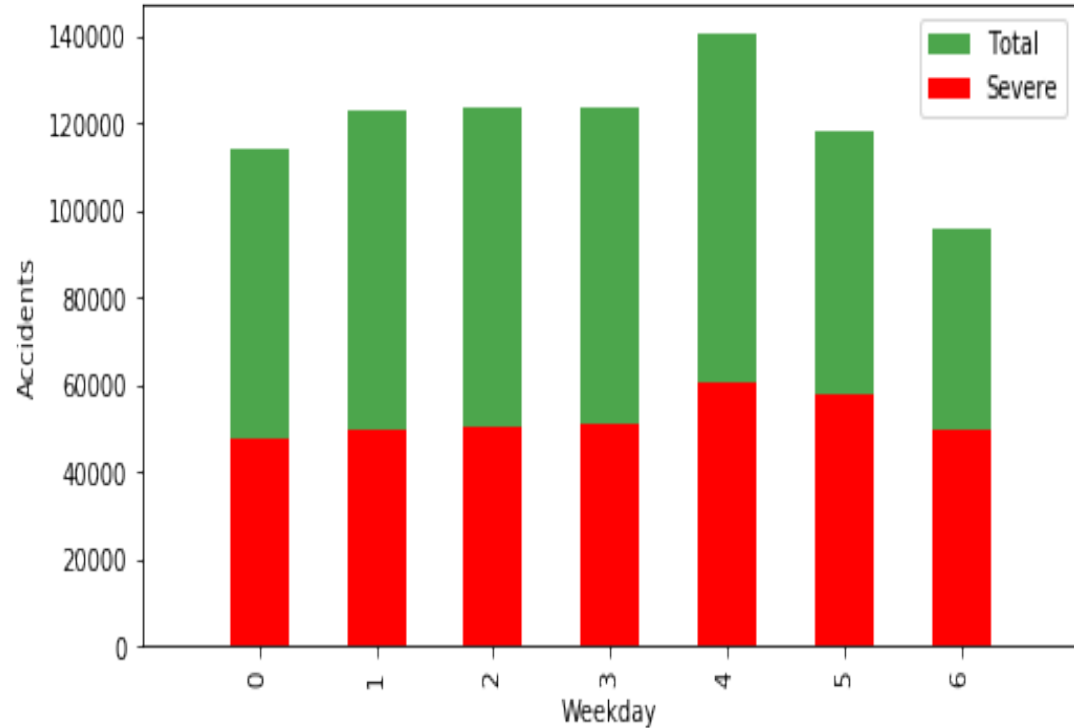


# EDA-Seasonality

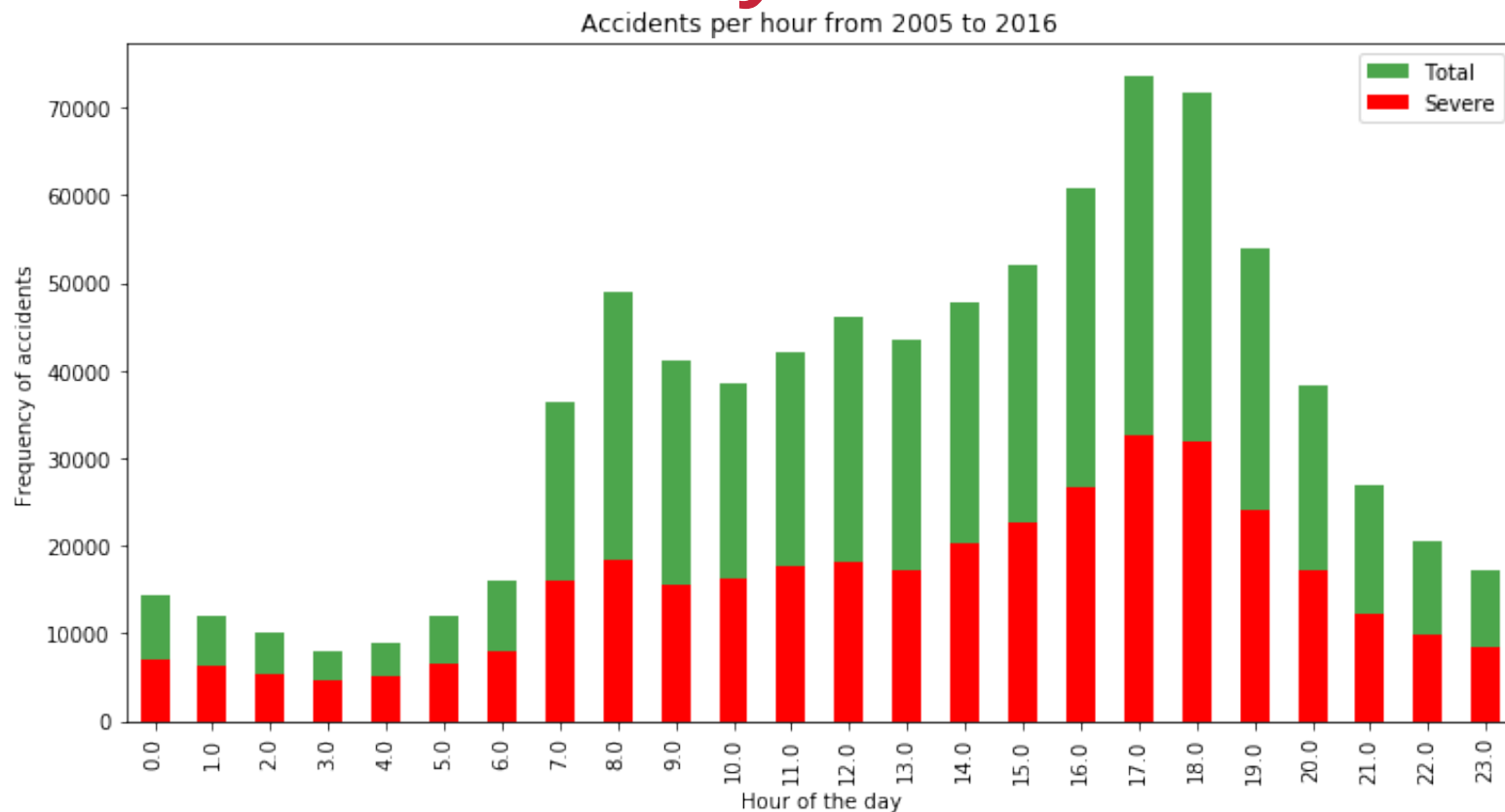
Accidents per month from 2005 to 2016



Accidents per day of the week from 2005 to 2016



# EDA-Seasonality



# Classification Models

## **Random Forest:**

- 10 decision trees
- maximum depth of 12 features

## **Logistic Regression**

- $C=0.001$

## **K-Nearest Neighbor**

- $K=16$

## **Supervised Vector Machine**

- Due to computation inefficiency, training size was reduced to 75,000 samples.



# Results

This table reports the results of the evaluation of each model.

Algorithm	Jaccard	f1-score	Precision	Recall	Time (s)
Random Forest	0.722	0.72	0.724	0.591	6.588
Logistic Regression	0.661	0.65	0.667	0.456	6.530
KNN	0.664	0.66	0.652	0.506	200.58
SVM	0.659	0.65	0.630	0.528	403.92

# Results

With no doubt the Random Forest is the best model, in the same time as the log. res. It improves the accuracy from 0.66 to 0.72 and the recall from 0.45 to 0.59.

