**Car accident severity**

Applied Data Science Capstone

AYUSH SOUN

<https://github.com/ayushsoun/Coursera_Capstone.git>

Traffic accidents are...

•Cause of 1.35milliondeathsgloballyin2016.

•Main cause of death among those aged 15–29 years

•Predicted to become the 7th leading cause of death by 2030.

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 to lives each year. Road safety should be a prior interest for governments, local authorities and private com- ponies investing in technologies that can help reduce accidents and improve overall driver safety.

All the recorded accidents in France from 2005 to 2016, both years included

•Initial dataset from the Kaggle, here.

•Preselected features on my GitHub, here

•In total 49 features, 839,985 rows in the Kaggle dataset

•Redundant and not relevant features were dropped

•29 features pre-selected

•On the data cleaning missing values and outliers were replace.

Road traffic injuries are currently estimated to be the eighth leading cause of death across all age groups globally, and are predicted to become the seventh leading cause of death by 2030.

Analysing a significant range of factors, including weather conditions, special events, roadworks, traffic jams among others, an accurate prediction of the severity of the accidents can be performed.

These insights, could allow law enforcement bodies to allocate their resources more effectively in advance of potential accidents, preventing when and where a severe accidents can occur as well as saving both, time and money. In addition, this knowledge of a severe accident situation can be warned to drivers so that they would drive more carefully or even change their route if it is possible or to hospital which could have set everything ready for a severe intervention in advance.

Governments should be highly interested in accurate predictions of the severity of an accident, in order to reduce the time of arrival and thus save a significant amount of people each year. Others interested could be private companies investing in technologies aiming to improve road safeness.

This project consist of several parts divided in two different notebooks.

## Feature Selection

This first notebook contains all the steps and transformations I performed for the feature selection. You can find the information on the raw data in the following [kaggle page](https://www.kaggle.com/ahmedlahlou/accidents-in-france-from-2005-to-2016). The kaggle datasets contain an extended descriptions of different aspect of the accidents, thus I've selected the most relevant and useful data for my analysis.

## **Predicting Traffic Accident Severity - Technical Overview**

* Data Description
* Data Cleaning
* EDA
* Model Development
  + Random Forest
  + Logistic Regression
  + KNN
  + SVM