Traffic Accident Severity Prediction

Abhijit Kumar Dutta

Background

- WHO research says 1.35 million road traffic deaths in 2016 with millions others suffering from severes injuries.
- Globally, traffic crashes are a leading cause of death among young people aged 15-27 years.
- Road traffic injuries are currently estimated to be the eighth leading cause of death across all age groups globally.
- Several factors contribute to the accidents. Few mentionable factors are weather conditions, road conditions and lighting, locatity and time of commute , type of vehicle etc.

Introduction

- The project aims to forecast the severity of accidents with previous information
- Data that might contribute to determining the likeliness of a potential accident occurring might include information on previous accidents such as road conditions, weather conditions, exact time and place of the accident, type of vehicles involved in the accident, information on the users involved in the accident and the severity of the accident.Road traffic injuries are currently estimated to be the eighth leading cause of death across all age groups globally.
- Governments should be highly interested in accurate predictions of the severity of an accident, in order to reduce the response time of the emergency medical services and to make a more efficient use of the resources.
- The prediction will also help private companies investing in technologies aiming to improve road safety procedures.

Data Acuisition and Cleaning

- Users, Places, Charateristics, vehicles and holidays data from kaggle dataset https://www.kaggle.com/ahmedlahlou/accidents-in-france-from-2005-to-2016

- 5 datasets were used consisting of all the recorded accidents in France from 2005-2016.
  1. Characteristics Dataset : Time, Place, Collision type, weather and lighting conditions.,etc
  2. Places Dataset : Road gradient, shape, category, surface condition and infrastructure,etc
  3. User Dataset : Place, user infomation, reason for travel, severity of accident, etc.
  4. Vehicle Data : Vehicle type
  5. Holiday : Labels the accident occuring on a holiday.

- Redundant and non relevant features are dropped, feature engineering is performed on the data set and finally 29 features are selected for model building.

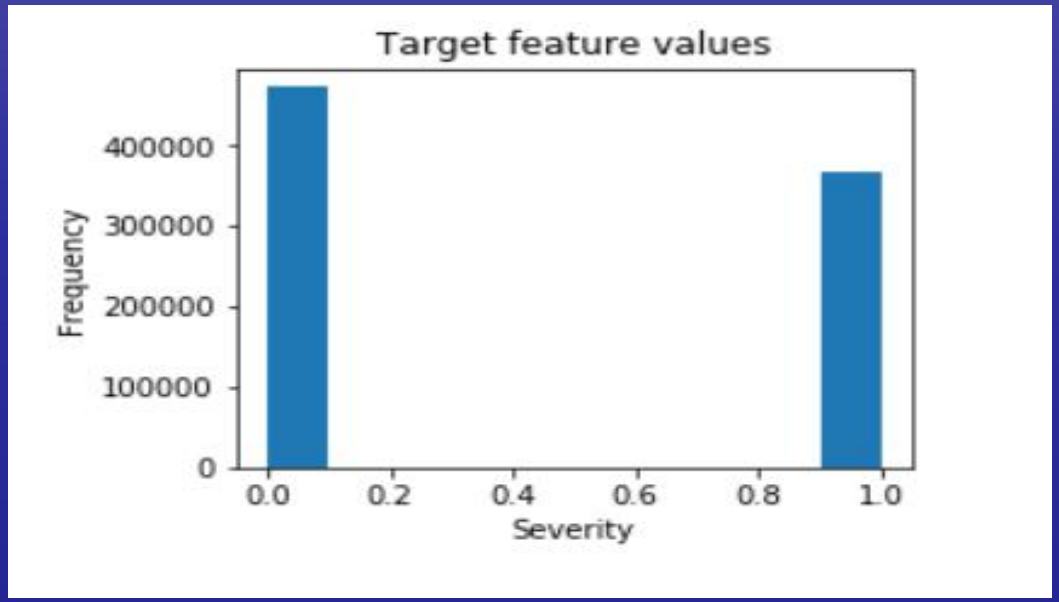
- Feature seection notebook link : https://github.com/abhijitcse414/Applied-Data-Science-Capstone/blob/master/Feature%20Selection.ipynb

Exploratory Data Analysis- Target

The target feature is a binary classifier depicting accident severity.

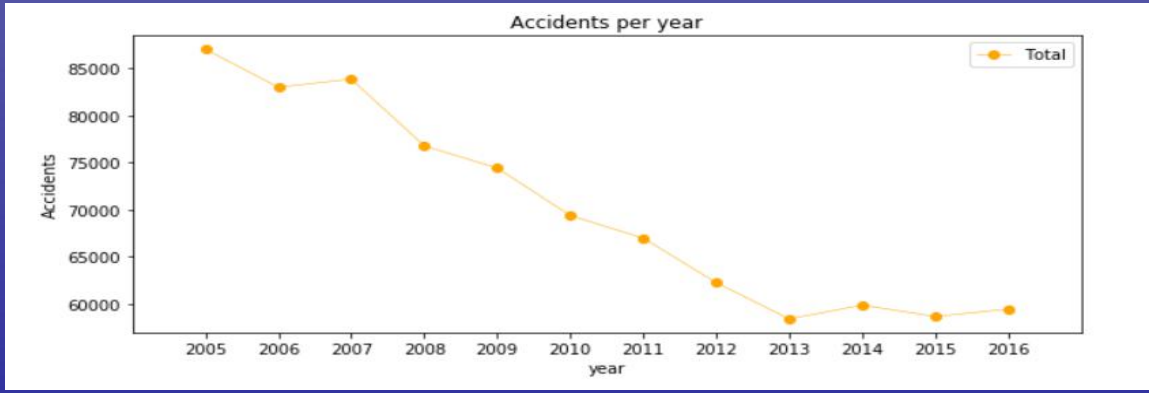
0 - Low severity

1 - High Severity i.e. hospitalized, wounder injuries to death.

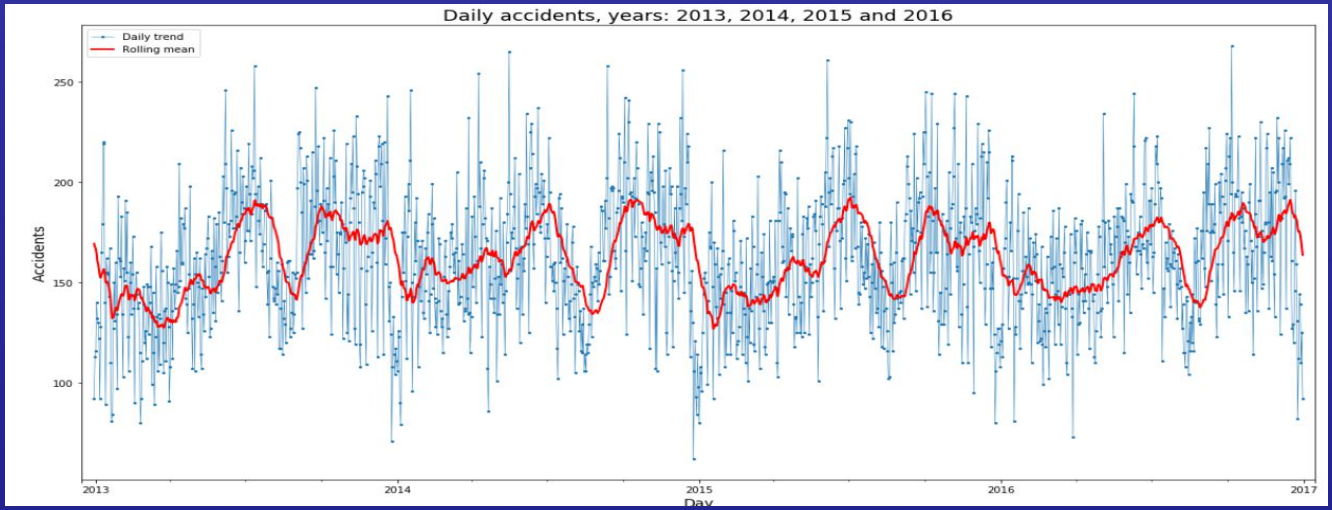


balanced label dataset with more low severity accidents.

Exploratory Data Analysis- Seasonality

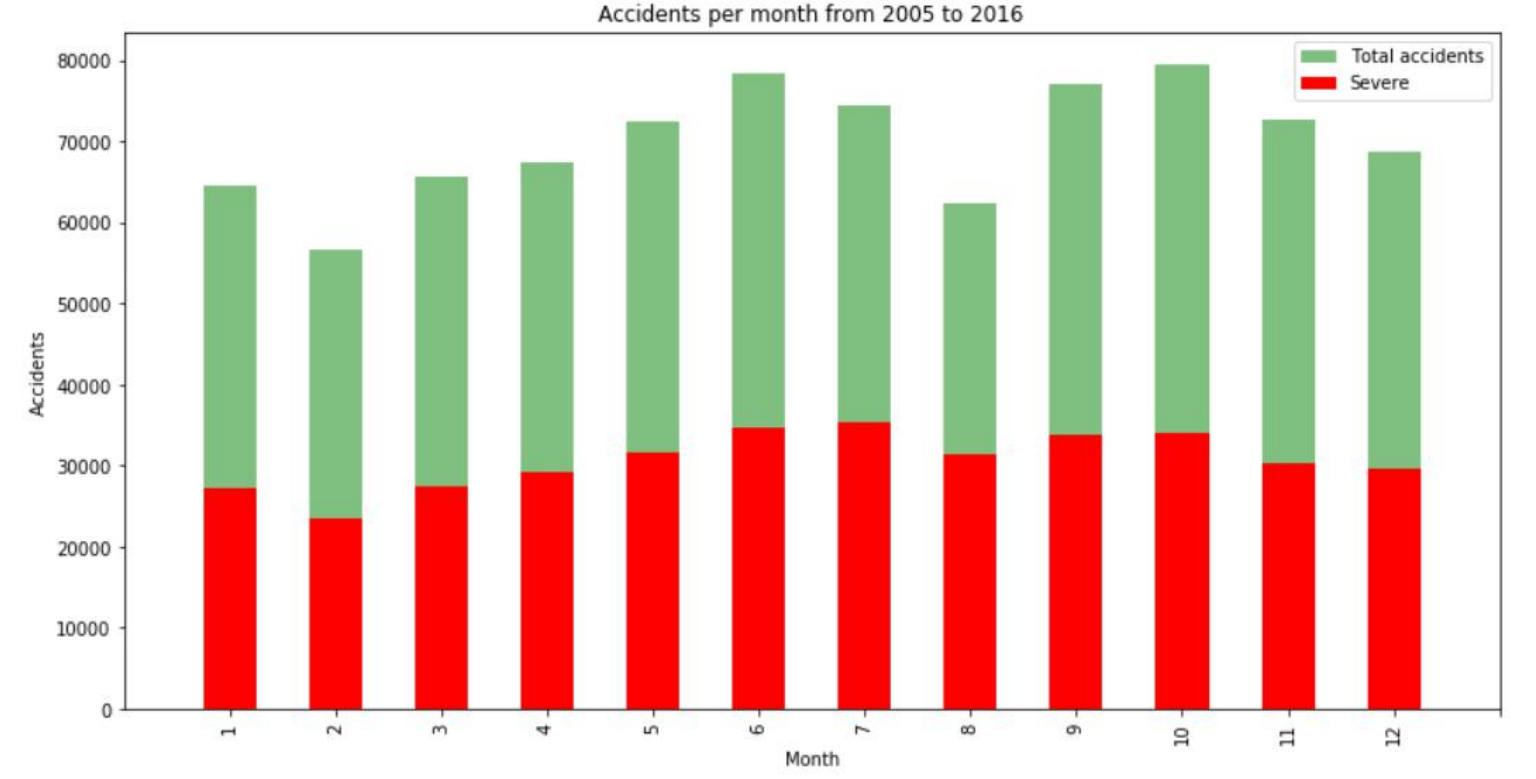


No. of accidents decreased from 2005 to 2013 and stabilised from 2013.



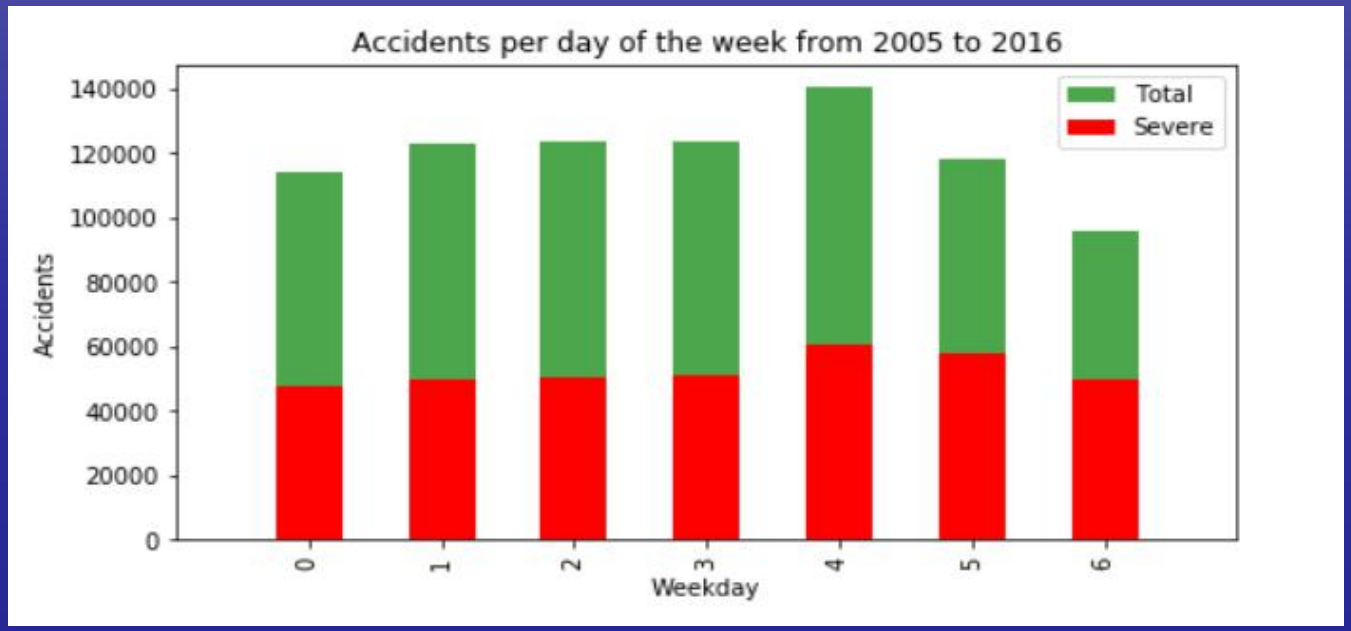
Exploratory Data Analysis- Seasonality

Accidents increase from March - June and with a spike in September and decreasing at the end of the year.



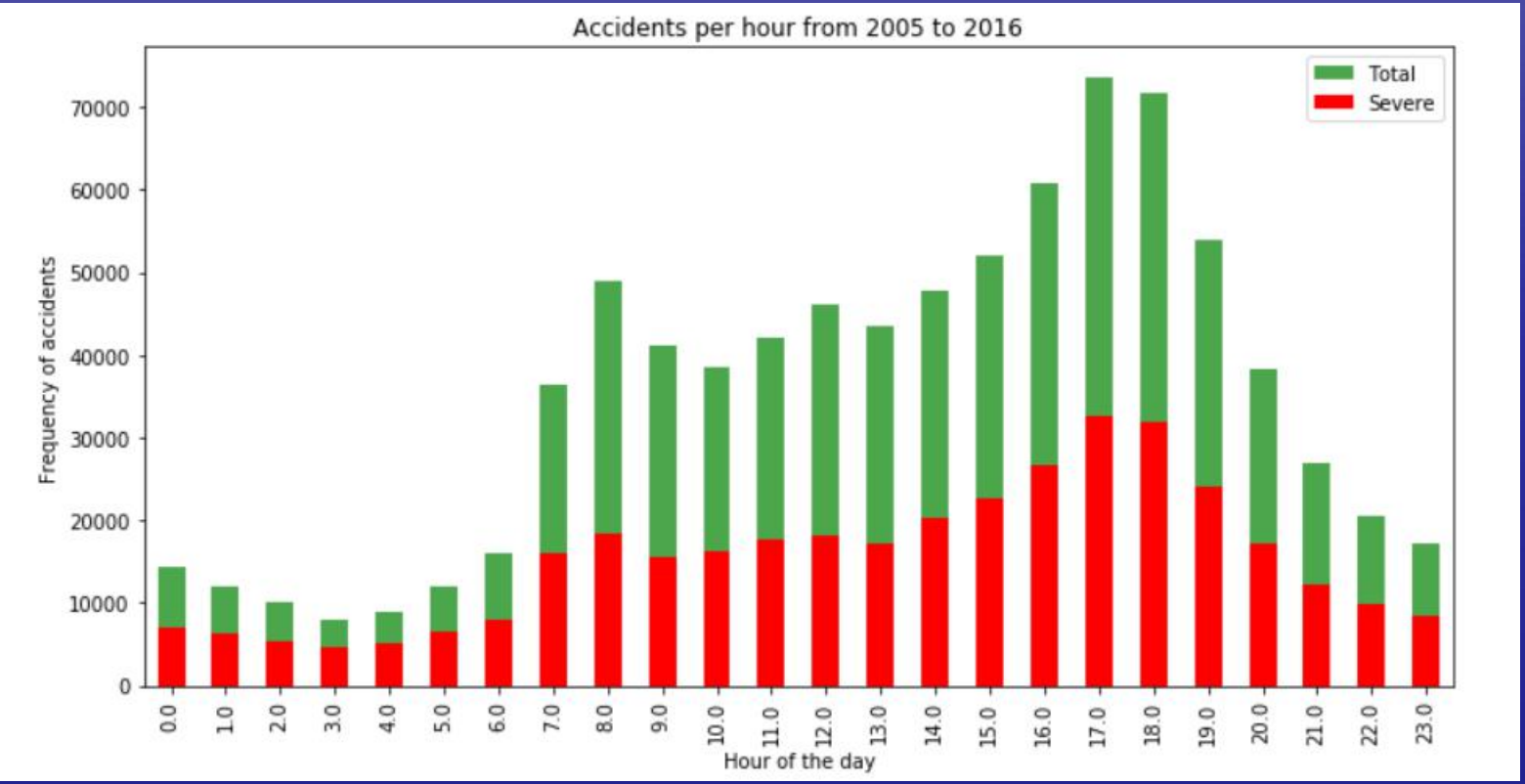
Exploratory Data Analysis- Seasonality

Steady Trend dyring the week with more accidents on Friday and less on Sunday.



Exploratory Data Analysis- Seasonality

Severe accidents occuring more during return from Office hours i.e. 5 p.m. - 6 p.m.



Model Building

Four Classification Models used for Model Building

Random Forest

10 Decision tress

Max depth - 12

Logistic regression

C = 0.001

K Nearest Neighbor

K = 16

Support Vector Machines

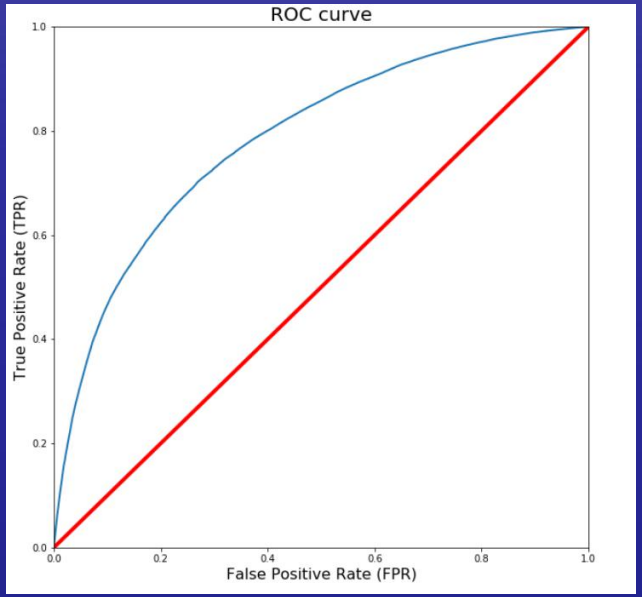
Training samples reduced to 75,000.

Results

Evaluation results of the Models.

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

Random Forest is the best model with
Accuracy - 0.72 and Recall - 0.59



Conclusion and Future Direction

Built useful models to predict whether and how much a player will improve.

Accuracy of the models has room for improvement.

Future Projects :

1. Add features such as vehicle and time of uninterrupted travelling.
2. Prediction of potential accident,critical spots and time.