Predicting Traffic Accident Severity

Geeta Akshata
October 2020

INTRODUCTION

- 54 million people worldwide sustained injuries from traffic collisions. This resulted in 1.4 million deaths in 2013, up from 1.1 million deaths in 1990.

- About 68,000 of these occurred in children less than five years old.Almost all high-income countries have decreasing death rates, while the majority of low-income countries have increasing death rates due to traffic collisions.

- The aim is to forecast the severity of accidents with previous information that could be given by a witness informing the emergency services.

Data

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

-Initial dataset from the Kaggle

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

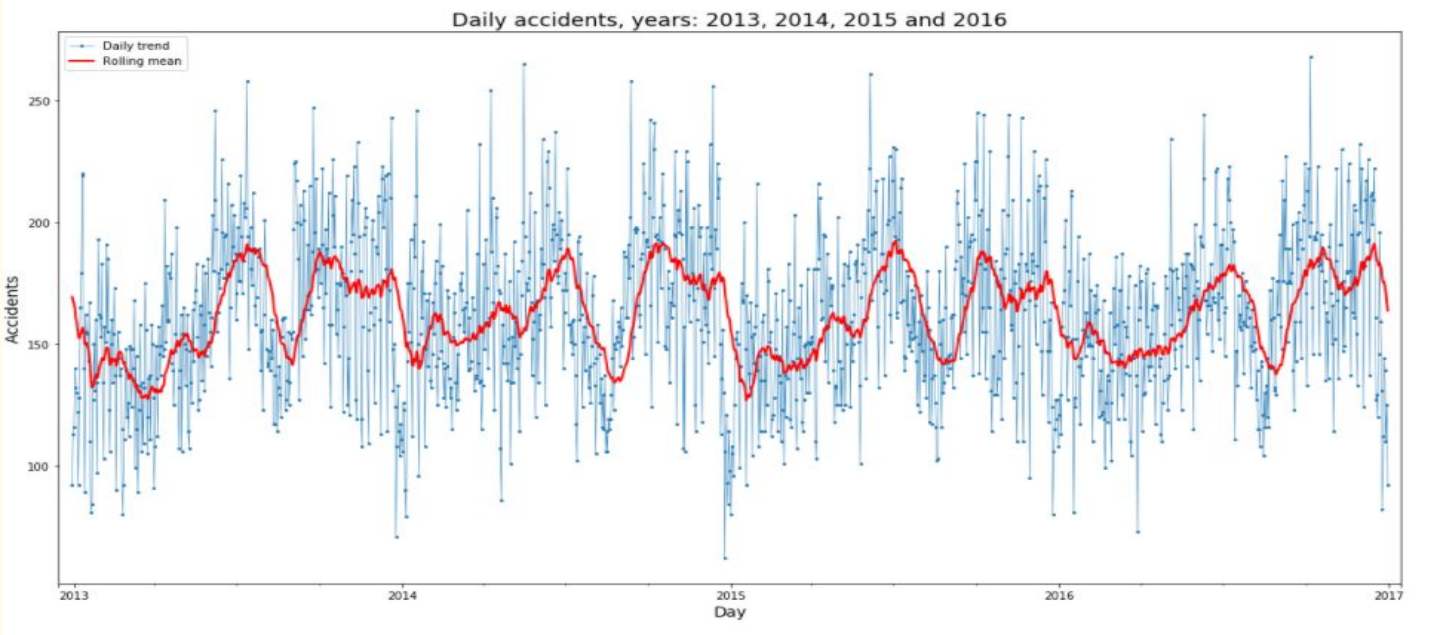
-Redundant and not relevant features were dropped

-29 features pre-selected

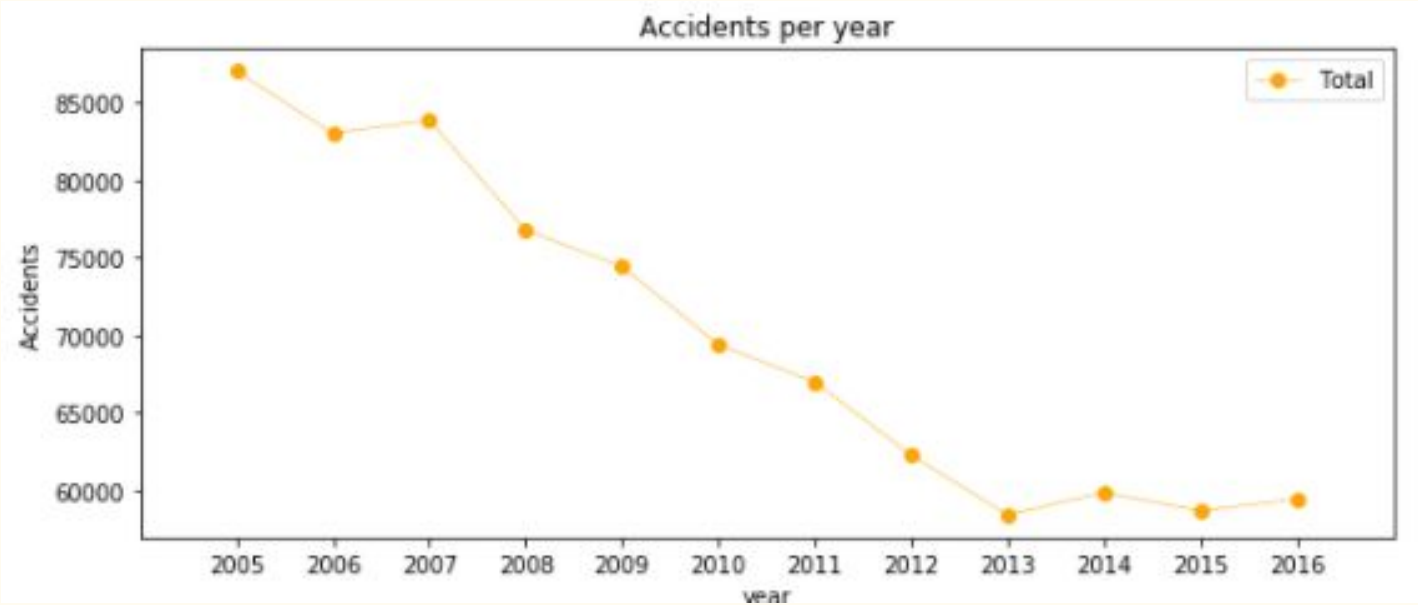
-On the data cleaning missing values and outliers were replaced.

Exploratory Analysis

- **Seasonality of accident:**



Accidents by year:



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 computational inefficiency, training size was reduced to 75,000 samples.

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.

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