UK Road Safety: Traffic Accidents and Vehicles

1. Introduction

The UK government collects and publishes (usually on an annual basis) detailed information about traffic accidents across the country. This information includes, but is not limited to, geographical locations, weather conditions, type of vehicles, number of casualties and vehicle manoeuvres, making this a very interesting and comprehensive dataset for analysis and research.

The data come from the Open Data website of the UK government, where they have been published by the Department of Transport.

The dataset comprises of two csv files:

- Accident_Information.csv: every line in the file represents a unique traffic accident (identified by the Accident_Index column), featuring various properties related to the accident as columns. Date range: 2005-2017
- Vehicle_Information.csv: every line in the file represents the involvement of a unique vehicle in a unique traffic accident, featuring various vehicle and passenger properties as columns. Date range: 2004-2016

Our target is to predict the accident severity. The severity is devided to two catagories; severe and slight.

We had more than 2 million observations and close to 60 features. So, we sampled the data into about 600K observations and 23 features.

Two models were selected - Logistic Regression and the Random Forest Classifier.

Context

The UK government collects and publishes (usually on an annual basis) detailed information about traffic accidents across the country. This information includes, but is not limited to, geographical

locations, weather conditions, type of vehicles, number of casualties and vehicle manoeuvres, making this a very interesting and comprehensive dataset for analysis and research.

The creation of this dataset was inspired by the one previously published by Dave Fisher-Hickey. However, this current dataset features the following significant improvements over its predecessor:

- It covers a wider date range of events.
- Most of the coded data variables have been transformed to textual strings using relevant lookup tables, enabling more efficient and "human-readable" analysis.
- It features detailed information about the vehicles involved in the accidents.

Content

The data come from the Open Data website of the UK government, where they have been published by the Department of Transport.

The dataset comprises of two csv files:

- Accident*Information.csv:* every line in the file represents a unique traffic accident (identified by the AccidentIndex column), featuring various properties related to the accident as columns. Date range: 2005-2017
- Vehicle_Information.csv: every line in the file represents the involvement of a unique vehicle in a unique traffic accident, featuring various vehicle and passenger properties as columns. Date range: 2004-2016

The two above-mentioned files/datasets can be linked through the unique traffic accident identifier (Accident_Index column).

The dataset will keep being updated as more data become available by the Department of Transport.

Acknowledgements

Thanks to Dave Fisher-Hickey for previously publishing, what I consider to be, the first solid and structured version of this dataset on Kaggle.

Also thanks to data.gov.uk for making this information publicly available.

Last but not least, thanks to The Data Lab for allocating me some much needed time to assemble this dataset.

Inspiration

Go crazy using the dataset. Don't go crazy while driving.

Datasource

Detailed dataset of road accidents and involved vehicles in the UK (2005-2016)

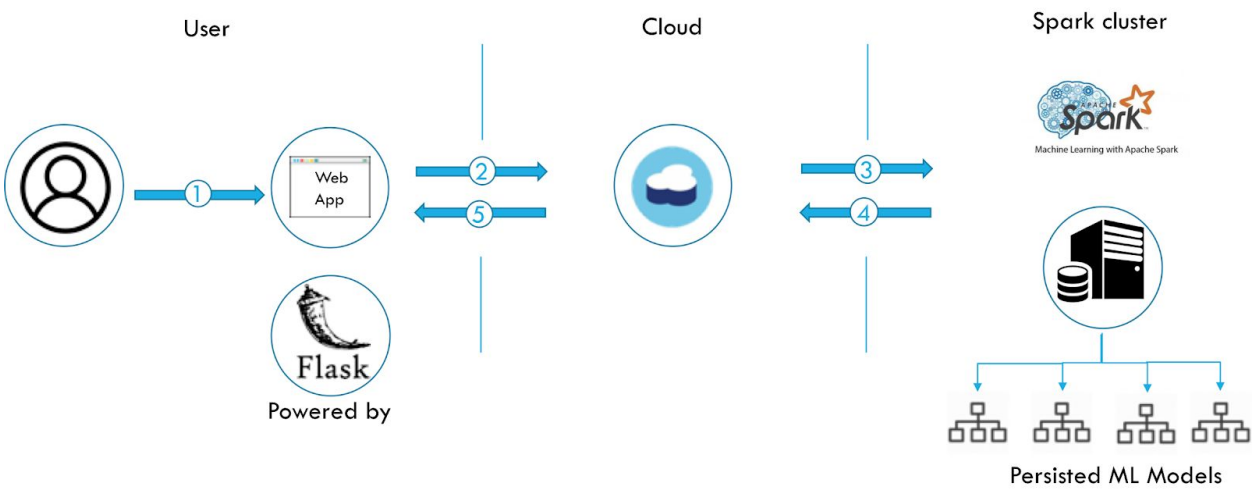
https://www.kaggle.com/tsiaras/uk-road-safety-accidents-and-vehicles#Accident_Information.csv

Use case

Predict accident severity at each output area (LSOA) given datetime and driving conditions For each Police Force we build a prediction model with the following objective:

Given date, time, weather, light and road conditions, predict accident severity within the operating geographic area of a police force

Project pipeline



Deliverable

The deliverable is a tool to faciliate ressource allocation and mitigate the probability of serious accidents given current driving conditions