**Predicting Accident Severity in Great Britain**

IBM Data Science Professional Certificate Capstone Project

Anne Bode, September 2020



**Introduction**

Certain everyday decisions come with a calculated risk, whether conscious or unconscious. It is relatively safe to say that the decision to get behind the wheel of a car and drive, alongside fellow cars, at speeds upwards of 60 mph comes with an inherent risk. However, more often than not we decide that the risk of getting into a severe accident is outweighed by the reward of getting to where we need to be. What if we could enhance our subconscious decision-making with a tool that would allow us to look out the window, contemplate our route, and decide we still like the odds? A tool like that could both save lives and improve traffic conditions, by reducing our likelihood of getting into a severe accident… and causing a bumper to bumper headache for fellow drivers. In this project, we look at accident severity data in Great Britain, specifically.

**Data**

The dataset used is provided by the UK’s Department of Transport and can be found [here](https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data). The “Road Safety Data – Accidents 2018” CSV file includes data on personal injury road accidents in Great Britain between 2005 and 2018. Additional information about the variables included in the dataset can be found [here](http://data.dft.gov.uk/road-accidents-safety-data/variable%20lookup.xls). Our goal is to use certain variables to predict whether an accident we might get in, over the course of our drive, would be “Fatal”, “Serious”, or “Slight”. Therefore, only a select number of variables within this dataset will be useful as predictive criteria. For example, latitude, longitude, and police force have to do with specific location and therefore would not allow us to create a generalized tool for the entire Great Britain population. Separately, a variable such as whether a police officer arrived at the scene would not help us create a predictive tool, as this is likely backwards looking, with police officers more likely to be on scene for more severe accidents. Consequently, we will restrict our dataset to just include features which an end user might reasonably be able to plug in *before* embarking on his or her trip, regardless of where in GB he or she may live. These include features such as the Day of the Week, Light/Weather/Road Conditions, and Urban or Rural area. We will use these features to predict our target variable: Accident Severity.