

Predicting Vehicle Accident Severity

An attempt was made to predict severity of UK vehicle accidents. Published UK Government national statistics data¹ was used, containing information of all reported UK vehicle accidents occurring on public roads in from Jan 1st 2016 to Dec 31st 2018. A total of 389,238 observations existed over this time period.

Information such as the number of vehicles and casualties, time and date, road type and surface conditions, speed limit, rural or urban area type, weather conditions and average age of casualties were used to predict severity. In the source database, accident severity was categorised into three levels: slight, severe and fatal.

Data from 2016 and 2017 were used to train models from various machine learning algorithms, which were tested in data from 2018. Performance was optimised through adjusting the parameters of the algorithms, and the algorithms were trained using equal numbers of observations for the three severity types to avoid bias from the unbalanced distribution of accident severity. A diagram of one decision tree model may be seen in the appendix, Figure 1.

Three out of the four models correctly classified fatal 2018 accidents with a precision of about 68-70%, each with good accuracy for prediction of slight accidents. One model behaved more poorly at predicting fatal accidents albeit achieving greater results for those with slight severity. However, all models struggled to distinguish between slight and severe but non-fatal accidents, resulting in poor performance for prediction of this category.

Exploring the dataset revealed interesting relationships between several variables and their impact on the severity of accidents. For example, it was seen that severe accidents occur relatively more frequently at night compared with non-fatal ones. Also, peaks of collisions are seen at 9am and 5pm on weekdays. Also, speed limit and average age of casualty had fairly strong relationships on severity outcome.

¹ Data.gov.uk. 2020. *Road Safety Data - Data.Gov.Uk*. [online] Available at: <https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>.

Appendix

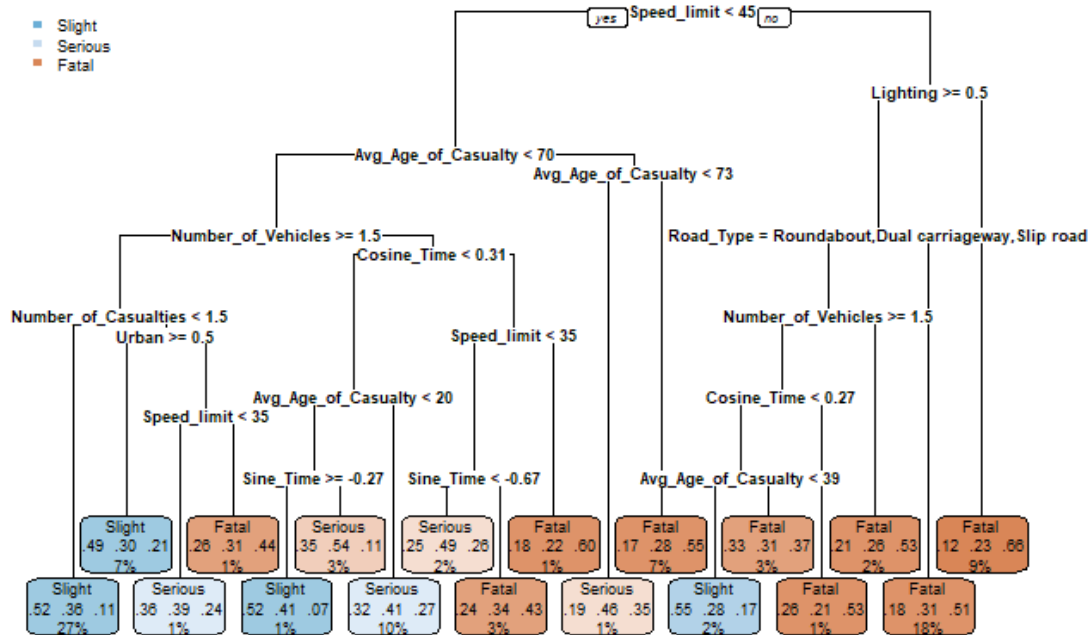


Figure 1: Example of a decision tree model developed. Categorisation of cases for the training data is displayed in the bottom boxes.

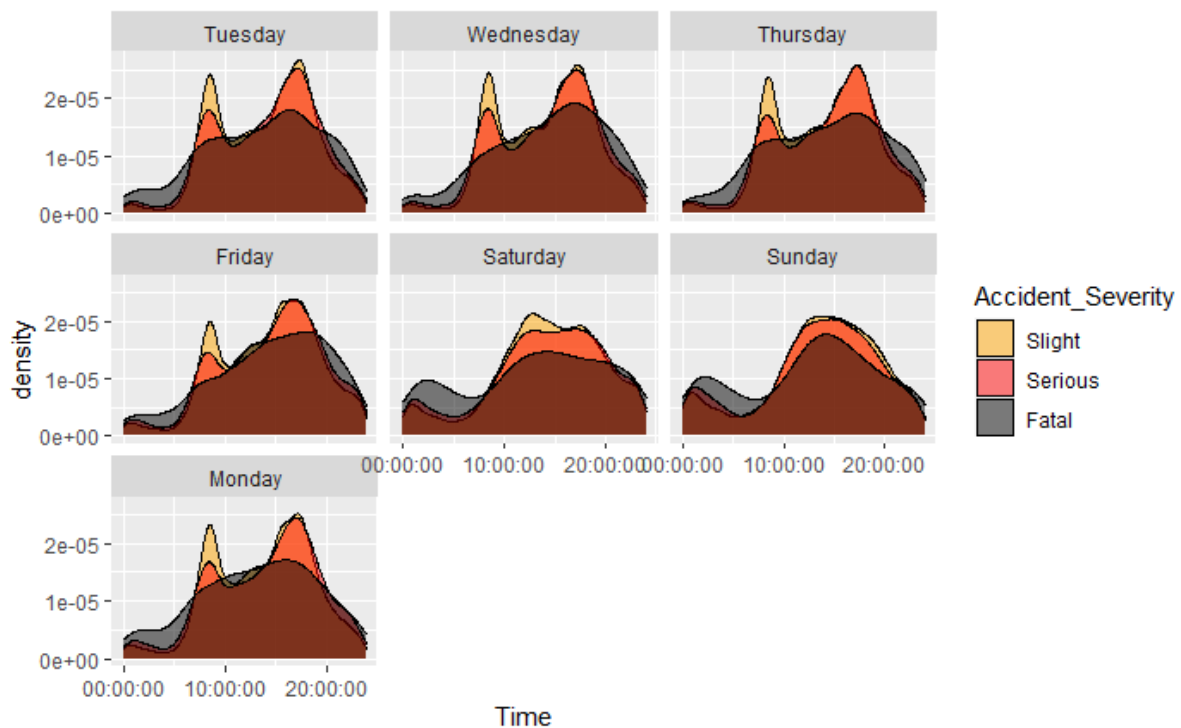


Figure 2: Accident severity distributions by time and day of week. Slight and severe accidents appear like each other when compared with fatal ones. Peaks are observed on weekdays at approximately 9am and 5pm and fatal accidents occur with greater relative frequency at night.