Coursera Capstone Project

Business Understanding

The objective of this capstone project is split into 2 parts:

1. Determine the probability of a traffic accident occurring given the conditions the car was driving in
2. If a traffic accident does occur, determine the severity of the accident given the conditions the car was driving in

This project could be helpful to multiple parties. Real-time mapping services could use this model to predict accident hotspots and give suitable recommendations to the drivers. Drivers who follow the recommendation could potentially avoid a traffic accident, and could potentially avoid possible property and human damage costs. The roadway authorities determine the accident blackspots in the area and make the necessary changes.

Data Understanding

Dataset is from the UK Open Data Platform and contains UK road accident data between 1979 and 2015. It was downloaded from kaagle and the link to the page is here:

<https://www.kaggle.com/akshay4/road-accidents-incidence>

The column 'accident severity' is the most suitable dependant variable, and it is a number from 1-3, where 1 is fatal, 2 is serious and 3 is slight

possible factors include

* 'Day of Week': days of the week represented by numbers 1-7. Weekdays and Weekends could be grouped together to see if more accidents occur on weekdays/weekends
* 'Road Type': Type of road where accident occured, e.g roundabout, dual carriageway, etc
* 'Junction Detail': Determines if there are a junction near the accident, and what type
* 'Junction Control': If there is a junction control, and what type
* 'Ped Cross - Human': If there is a specialized crossing nearby for schools etc
* 'Ped Cross - Physical': If there is a crossing nearby
* 'Light Conditions': Lighting condition
* 'Weather': Weather conditions
* 'Road Surface': Road Surface
* 'Special Conditions at Site': If there are special obstacles in the area e.g roadworks
* 'Carriageway Hazards': If there are any obstacles on the road e.g other vehicles, dogs, etc
* 'Urban Rural': Determines whether the accident happened in an urban enviroment or a rural enviroment
* 'Journey Purpose': Purpose of trip
* 'Age of Driver': determines age of driver

All the data columns are categorical apart from the age of the driver. That column of will have to be normalized. Certain dummy variables of the data columns would have to be sorted again (e.g grouping weekdays and weekends for 'day of week').