COURSERA – CAPSTONE PROJECT

Car accident severity

**BUSINESS UNDERSTANDING**

Car accident data and machine learning can be used to give drivers a better information about the risk to have a severe car accident in a certain area and under certain conditions like weather, road and visibility. Applications like Waze, Google Maps or Tom Tom GPS Navigation Traffic could usefully display this information to the drivers when crossing the area.

The information will be based on a model predicting the severity of road accidents under conditions. All the data come from the dataset provided by the SDOT Traffic Management Division in Seattle. Once the model completed and the prediction available, real time conditions (weather, road, and visibility) could be added to give a more precise information to the driver.

**DATA UNDERSTANDING**

The dataset from the SDOT Traffic Management Division in Seattle presents a part of the car accidents and their characteristics detailed in 37 columns. All the columns are labelled but features are not always completed. Only the rows having a full set of data corresponding to the selected features will be used for the model. The column OBJECTID will be the unique identifier of each row.

Attributes needed to build the model:

Independent variables: X; Y; ADDRTYPE; LOCATION; WEATHER; ROADCON; LIGHTCOND

Dependent variable: SEVERITYCODE

Data must be prepared by converting it to numerical type (ex: INT for integer); some rows containing missing data must be removed.

Finally, data must be balanced in order to achieve an equal representation of both SEVERITYCODES. Therefor evaluation metrics for classifiers (like F1-score) will be used.