COURSERA – CAPSTONE PROJECT

Car accident severity

**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.