Capstone Project Report –

Severity Index of Accident

Introduction | Business Understanding

This project work deals with an attempt to reduce the number of vehicular accidents/collisions in a community. A machine learning prediction algorithm is to be developed to predict the severity index of an accidental scenario due to the conditions like the current weather, road and visibility conditions on a particular prone area. When conditions are bad, this model will alert drivers, travelling on that particular route, and remind them to be more careful.

Data Understanding

Here, we are using the dataset available on the course website itself.

The Dataset deals with the Feature called “SEVERITYCODE”. It is a categorical target variable which discusses the damage index. ‘1’ means Low Change of Risk or “Property Damage only” type of collision & ‘2’ means High Change of Risk or possibility of “Injury”.

This target variable is wholly based on some input attributes. These inputs are “WEATHER”, “ROADCOND” and “LIGHTCOND”. These features discuss the conditions that the driver faces. Namely weather conditions, light and visibility during the accident and the condition on the road.

As the data is not fit to be used in the model analysis and training in its current form, thus, we extract our useful features and discard many of the features that are not being used.

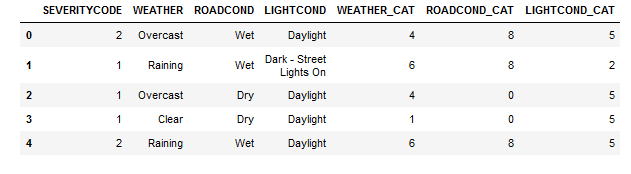


Figure 1: Sample dataset with extracted featuers

Then, we labelled the categorical attributes into our desired numbered form using LabelEncoder.