

Applied Data Science Courser Project Report

Introduction/Business Problem

Accidents in the US and around the world are a significant problem that increases the fatality rate of a country and also leads to economic loss. With the improvement in machine learning over recent years, it is essential that we study the effects of various attributes on roads and can effectively model them allowing us to predict accident severity. The input to such a model would be the characteristics of previous accidents such as driver behavior, road conditions, road type, environmental conditions, and the output vector would be the class of accident severity. By building such a model, we would enable the drivers to have information about the seriousness of getting into a road accident if they kept driving on the road that they are on. This extra piece of information could make sure that the driver changes their style of driving on a particular route or change their course to avoid accidents. Therefore, it is essential to invest in building a road accident severity prediction model.

Data

The Data set that I will be using for this project is obtained from coursera.org for Seattle city. This dataset has 194673 rows and 37 attributes about the road. Attributes such as Location, Road condition, weather condition, junction junction, car speeding, number of people involved, light conditions, number of vehicles involved in are some of the attributes that can be used to create the model. The label of the Data is present in the column Severity. There are unbalanced labels present in the dataset which will need to be balanced in order to reduce the bias in the model.