Accidents have been a major contributor in the number of deaths per year especially in densely populated cities. More severe the accident, more are the chances of the person dying. What can we do to prevent the loss of lives in accidents? If, by the means of data analysis, we are able to create a model that would predict the degree of severity of an accident by taking into factor certain ‘variables’ which can influence the severity, for instance, the severity factor of an accident might increase in night-time as opposed to that in daytime, then we can be prepared beforehand by knowing probability of the degree of severity based on the data for that particular area. The example dataset that is provided to us consists of such variables that can help predicting the severity of accident. In simple words by knowing the conditions of traffic of a particular area, the model will predict the severity of an accident.

Since the output, i.e., the values of the degree of severity that we want is divided into two classes (‘1’ and ‘2’), this problem is a classification one. We have to create a model which will take in a certain number of independent variables and predict either of the class. Thus, the problem of predicting the degree of severity is a classification one and since there is only a single dataset, I will split it into the training and testing sets and will then proceed to measure the accuracy of different models. The model with the highest accuracy will be selected for the purpose of prediction.