A description of the problem and a discussion of the background.

Reducing traffic accidents is an important public safety challenge, therefore, accident analysis and prediction has been a topic of much research over the past few decades. The problem is to identify the possibility of getting in a car accident based on several different factors as given in the shared dataset. Say you are driving to another city for work or to visit some friends. It is rainy and windy, and on the way, you come across a terrible traffic jam on the other side of the highway. Long lines of cars barely moving. As you keep driving, police car start appearing from afar shutting down the highway. Oh, it is an accident and there's a helicopter transporting the ones involved in the crash to the nearest hospital. They must be in critical condition for all of this to be happening. Now, wouldn't it be great if there is something in place that could warn you, given the weather and the road conditions about the possibility of you getting into a car accident and how severe it would be, so that you would drive more carefully or even change your travel if you are able to.

In the shared dataset, there are 37 attributes (columns). The first column tells the severity of the accident. The remaining columns have different types of attributes. Some will be used to train the model. The label for the data set is severity, which describes the fatality of an accident. Some of the shared data has unbalanced labels which may create unbiased ML model so that will require balancing the data.

Target audience

- Vehicle Drivers
- Ambulance services

The drivers would care because they would not want to get in an accident. Ambulance services can use the model to predict accident severity and be ready to respond in a timely fashion.