**Predicting the Severity of Accidents in Seattle**

Manohar Emani

August 30, 2020

1. **Introduction:**

Road accidents have always been an unsolved puzzle in the history of transportation. Road accidents have bad effects on various areas of life. A road accident can lead to physical damage to the victims and at times, it can lead to loss of lives. In the business perspective, road accidents result in damage of goods and delays in delivery of services. For the government, road accidents lead to huge traffic blocks. Public properties like the roads, pavements, etc., also get damaged due to the road accidents. To avoid more accidents, it will be helpful if the accident severity at various locations of the city of Seattle can be predicted based on the previous data so that the travelers can avoid the locations or at least, take care while they drive through the locations with high severity of accidents.

All the stake holders including the travelers, the government and the people who live near those locations of Seattle city show their interest to know the predicted severity of accidents so that they would cooperate with each other to resolve the issues in locations according to the predicted severity of accidents and make them safer for travel. Government can plan new infrastructure in those locations to avoid accidents.

1. **Data:**

The dataset provided in the Capstone course is used to build a good machine learning model that can predict the accidents’ severity of a location based on various attributes. The data is downloaded from [here](https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv). The data has 37 attributes like the location, junction type, weather condition, light condition, accident severity, etc., and 194,674 records of various accidents happened at various locations in Seattle city. The data contains redundant columns, missing data, etc. All these anomalies are fixed and a few columns that are highly related to the accident severity are selected. Then, the processed data is used to build a good model that can predict the SEVERITYCODE (the column that contains accident severity codes based on the type of severity) based on the columns that are selected for the analysis.