Predicting Car Accident Severity

Peiming Wu

August 21,2020

1.Introduction/Business Problem

1.1 Background

There are numerous car accidents happen across the world every day. All car accidents will create damage to cars involved or even worse, take lives. People want to drive safe not only in terms of reducing the chance of damaging their cars but also less life risks. In fact, there are many factors that contribute to the severity of a car accident. Therefore, it is advantageous for related departments to accurately predict the severity of car accidents under those conditions. For example, does bad road conditions involves in large number of car accidents? If it does, the prediction provides solid reason for better road constructions. From that, warning information like road signs to the drivers will lead to positive impacts.

1.2 Business Problem

Data that might contribute to a car accident including locations, weathers, road conditions, light conditions, vehicles or pedestrians involved, speeding, whether the driver is involved was under the influence of drugs or alcohol, etc. This project aims to predict the car accident severity based on these data.

2.Data acquisition and cleaning

2.1 Data Sources

The dataset that will be employed in this project is the example dataset provided by the teaching staff. It includes all collisions provided by SPD and recorded by Traffic Records. The dataset could be downloaded from https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv and the Metadata could be downloaded from https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf . This dataset is a supervised with labeled severity of car collisions with numerous attributes like locations, weathers, road conditions, light conditions, etc. This dataset, however, needs to clean since there are empty inputs in some attributes like road conditions.