

Predicting Car Accident Severity

Introduction

There are 280 million cars in the United States [1]. Such high level of traffic causes many accidents. In 2018, there were 12 million car accidents causing 36,560 deaths [1]. Several factors cause car accidents, such distracted driving, speeding, weather, drunk driving, etc.

To address this major problem, a machine learning model maybe developed to predict the severity of car accidents. Predicted car accident severity could help to warn drivers of potential risks and improve the preparedness of first responders.

Data

The data used in this project includes all types of collisions in Seattle from 2004 until 2020, with a total of 194,673 accidents. The data includes several attributes, such as accident location, address type (ally, block, or intersection), weather, road condition, speed, lane segment, etc. Attributes that are related to the causes of accidents, such as speeding and road conditions will be examined as they may help to predict the severity of an accident.

The dependent variable which is the severity of an accident has four levels:

- 0 refers to unknown.
- 1 refers to accidents with property damage.
- 2 refers to accidents with injuries.
- 2b refers to accidents with serious injuries.
- 3 refers to accidents with fatalities.

For this project, accident with severity level 0 (unknown) will be deleted as they do not contribute in training or testing our model to predict the severity of car accidents. To deal with imbalanced data, data will be under-sampled to balance the different level of accident severity.

References

1. <https://www.statista.com/topics/3708/road-accidents-in-the-us/>