Accident Severity Prediction

September 10,2020

1. Introduction

We will be using data provided by the Washington State Department of Transportation Crash Data Portal. We will investigate the severity of car accidents caused by bad weather. The data used are such as: weather, location, and road condition provided in the dataset, which is available from the Seattle Department of Transportation's Division of Traffic Management. With the conclusions we draw, our goal is to prevent or reduce serious or fatal accidents in the future.

2. Data Understanding

With the data provided by the Seattle Department of Transportation Traffic Management Division, accident information will be covered from January 2004 to May 2020. The data is segregated into location, weather condition, type of accident, status, severity. There are 194,673 observations and 38 variables in this data set. Since we would like to identify the factors that cause the accident and the level of severity, we will use SEVERITYCODE as our dependent variable Y, and try different combinations of independent variables X to get the result.

3. Methodology

We used Jupyter Notebook to do the data analysis. To generate the table and graph for the dataset, we imported Python libraries (Pandas, Numpy, Matplotlib, and Seaborn).

Columns with data not relevant to the investigation were eliminated. These include: "OBJECTID", "INCKEY", "REPORTNO", "COLDETKEY", among others. Cells with unidentified data were identified. These were replaced with a N / A in order to establish a uniform data.

```
df.replace(r'^\s*$', np.nan, regex=True)
df.replace("Unknown", np.nan, inplace = True)
df.replace("Other", np.nan, inplace = True)
```

Then the columns "INATTENTIONIND", "PEDROWNOTGRNT", "SPEEDING", were removed as data was missing. Applying this decreased the data from 19,4673 rows to 14,747 rows.