Introduction

Besides causing injuries and deaths, traffic accidents also result in significant property and economic losses. In the United States, 40,000 lives are lost annually, with an estimate economic losses of $80 billion. Therefore, this research aims to analyze the various factors that could help predict accident severity can guide the board to implement changes to reduce traffic fatalities in a timely manner, using various analytical techniques and machine learning algorithms.

Business Problem

The objective is to analyze the collision data set for Seattle, WA and determine which is the most significant factor that causes traffic accidents such as weather, road conditions, visibility and how can we predict its severity. The relevant stakeholders can thereafter make informed recommendations based on these findings to formulate and implement policies to alleviate this issue.

Target Audience

The intended audience for this project will be the Department of Transportation of Seattle, Washington so that they can use the findings to make recommendations for its citizens.

Data

The data is based the statistics provided and compiled by the Seattle Department of Transportation from 2004 to present. The information includes the various factors and conditions in which the accidents occurred with a total entry of 194,673.

It also contains a severity code that ranges from 0 (unknown) to 3 (fatality). Therefore, it is able to predict the severity level of an accident based on these features in the dataset.

The following will be done for data treatment: Data cleaning by removing the NaN values, correcting the data types and dropping any outliers. Modelling by train and measure models. Logistic regression, linear svm classifier, decision tree and random forest classifier to train a baseline model and tuning a better model will be done.