Seattle Traffic Accident Severity Prediction

Guankai Zhai

August 20, 2020

I. Introduction

1. Background

According to the US Department of Motor Vehicles (DMV), about 6 million car accidents happen each year in the US. On an average day, over 90 people die due to traffic accidents, leaving hundreds of families behind them. Despite such disturbing facts, people's awareness of traffic safety is still insufficient. According to a survey, only 1 in 7 people will wire a seatbelt when he or she sits in a car, while seatbelts could reduce the risk of death by over 45%. Therefore, it is advantageous for people to realize the risk of traffic accidents and thus practice safer driving habits.

2. Problem

This project aims to determine which factors affect the severity of car accidents and then build a model that could predict the severity of car accidents based on the determined factors.

3. Interest

The information from this project would be beneficial to many parties including the DMV, police departments around the country, and most importantly, drivers and their passengers. The project would shed light on what combination of conditions are conducive to fatal car accidents, how the severity of car accidents depend on these conditions, etc. After knowing the relevant information for a given road on a given day, the police department could predict the possibility of fatal car accidents and therefore warn drivers to practice safe driving styles.

II. Data

1. Sources

The dataset that this project will use is provided by the International Business Machines (IBM) Corporation here.

2. Content

The dataset has 194672 instances of traffic accidents that occurred in Seattle. It labels each collision as either of severity code 1 or 2, representing prop damage or injury. Complete metadata of this dataset could be found here.