Predicting Car Accident Severity in Seattle
Week 1

Ana Flávia Santos Souza

1. Introduction

In the US, there are around 6 million car accidents every year, and 90 people die because of that every day[1]. Great centers in the country, such as Los Angeles, New Orleans, and Baltimore, are usually the ones where most accidents occur[2]. Seattle, however, appears to be in the opposite direction. Even though it is one of the biggest cities in the country, when it comes to driving, Seattle is considered one of the safest places in the state of Washington[3]. That fact doesn't diminish the need to continually evaluate the city's accident rates and keep it track of how much damage each of these collisions causes, especially to guarantee that the rates will remain low.

Fortunately, data have become accessible as well as methods for visualization and analysis, making it easy to evaluate these rates. With the right data, it is possible to evaluate, identify, and predict how harmful an accident was or will be, using machine learning algorithms. This allows authorities to know how much services and efforts should be designated for this issue.

1.1. Problem

The analysis of accident rates is important to understand how much damage is caused by a collision, and what can be done to decrease the harms. This project aims to use machine learning algorithms to predict car accident severity based on data of a period of fifteen years (2004-2019) from Seattle, WA.

1.2. Target Public

The key point of this project is to predict car accident severity based on the accident characteristics. This could be useful for the authorities responsible for assisting in these cases, such as police and medical services, to better understand how severe an accident is before getting into the location.

---

[1] https://www.driverknowledge.com/car-accident-statistics/
[2] https://quotewizard.com/news/posts/americas-most-car-accident-prone-cities
[3] https://quotewizard.com/news/posts/washington-most-distracted-driving-cities

2. Data Collection

To examine car accident severity in Seattle, I used a CSV file provided by the Applied Data Science Capstone course on Coursera. The course also offered a metadata file with a description of the columns and their types. The original dataset was from the Seattle Police Department and Accident Traffic Records Department. It had 38 columns containing information about car accident occurrences from 2004 to 2020, including date and time of the accident, the accident's location, type of collision, number of people and vehicles involved, accident severity, and so on. With this data, I expected to get a perspective about the reasons that may lead to a car accident and predict the accident severity.