**Predicting Car Accident Severity**

IBM Data Science

Capstone Coursera Project

Author: Noor Kouli

Data of Submission: 10/09/2020

Abstract

Table of Contents

[Introduction 4](#_Toc50641546)

[Data 5](#_Toc50641547)

# Introduction

Car accidents are a common problem that occur daily. They can result from speeding, tailgating, failing to use an indicator, bad weather conditions or any form of distraction. However, some accidents are more severe than others depending on the factors mentioned earlier that can lead to permanent injuries or even death. In addition, car accidents can cause heavy traffic and may block roads causing people to miss important meetings, flights and so on. Given the weather and road conditions, what if we can predict the possibility of getting into a car accident and how severe it might be? This will signal us to drive more carefully, change our route or postpone a meeting ahead of time. Not only will this solve the problem of wasted time but will also enhance awareness of our surroundings to safely arrive at our destination.

In the past, researchers and data scientists have built machine and deep learning models on different car accident scenarios since it is a very common problem they are trying to solve. For example, the automatic detection of a car accident from the driver’s phone in order to receive medical care in a timely response. This detection can help save time and allow the ambulance to be on its way as soon as the accident occurs. Another example is detecting human driver inattentiveness through capturing and analyzing the face of a driver. In addition, it can detect fatigue, drowsiness and aggressive behavior. Once detected, this deep learning algorithm with signal them to focus on the road or stop on the side.

Although there have been quite a few studies on car accidents, there is still ongoing research on how extreme weather conditions and traffic affect car accident severity. Therefore, this project encompasses a machine learning model that takes these specific factors into account to help minimize car accidents overall.

# Data

The data frame for this model consists of 146,059 rows and 38 columns that address collision history in Seattle, Washington from 2004 to the present day. The columns are collision type, location of accident, junction, number of people involved in the accident and so on. However, not all the 38 columns in the dataset can be used. Therefore, this dataset must be cleaned in order to begin building the model. The primary column in this dataset that we are focused on is the severity code. This severity code can be values of 0,1,2,2b and 3. 0 meaning the severity is unknown, 1 is property damage, 2 is injury, 2b is serious injury and 3 is a fatality. In addition to the severity code column, the address type, location, collision type, vehicle count involved in the collision, weather conditions, and whether the driver was speeding or not will be the variables involved in classifying a test set of car accidents into its respective severity code. This dataset will be divided 80% into a training set and 20% into a test set in order to achieve out of sample accuracy. The test set will help us classify accidents into a severity code number and help the driver change his route or slow down his pace depending on the value of the severity code.