

# *CAR ACCIDENT SEVERITY*

CAPSTONE PROJECT

*Joseph Zahar*

*IBM- Applied Data Science*

# 1 Introduction

All around the world, roads are shared by many motorized vehicles that have made transportation faster and more comfortable while supporting many countries' economic and social development. However, these vehicles cause a global problem. Car accidents are responsible for 1.35 million deaths on roadways every year. Almost 3,700 people are killed globally in road traffic crashes, where more than half of those killed are pedestrians, motorcyclists, and cyclists. Road traffic injuries are estimated to be the eighth leading cause of death globally for all age groups and the leading cause of death for children and young people 5–29 years of age [1]. Car accident severity depends on many factors and usually causes a high volume of traffic.

The main business problem is the severity of these accidents, which can sometimes be fatal and critical for pedestrians, bicycles, or vehicles. The project's objective here is to get people to safety, and the best way to solve this problem is to prevent those accidents as accurately as possible. Imagine being able to predict in advance the probability of an accident happening depending on the weather, the type of road, and many other features! This would help save many lives and help drivers get to their destination in the safest and fastest route possible.

The audience involved for this kind of project would be any Weather program, News channel, Radio station, government, or mobile application developer that wants to inform drivers of the possible areas of road crashes in a community and propose alternative solutions to get to their destination.

## 2 Data

For this project, the Dataset was shared on Coursera as a csv file available by [Clicking here](#). The data is composed of 37 features that accurately describe each car accident that happened in Seattle from 2014 to 2020. They are classified in terms of severity, type of weather and road condition, location, address type, and many more. However, for this case study, not all attributes are useful as the main objective is to predict an accident's probability and severity. Therefore, the Dataset needs deep understanding and analysis before choosing the right attributes to reach our goal. After selecting the appropriate features, the new Dataset is balanced and preprocessed before feeding it to a supervised machine learning model that will learn to predict in the future the probability of a car accident based on the chosen attributes, in the most accurate way.

To consider the problem we can list the data as below:

- After visualizing and comparing the effect and weight of many attributes on the severity of an accident, the main attributes that will be used are the following:

- 'WEATHER': A description of the weather conditions during the time of the collision.
  - 'ROADCOND': The condition of the road during the collision.
  - 'LIGHTCOND': The light conditions during the collision.
  - 'LOCATION': The location of the accident.
  - 'SEVERITYCODE': Representing the severity of each accident
- After getting the number of labels for each type, we can clearly see that the data is unbalanced as there is a lot more of accidents of type 1 than 2. To balance the data we may either down sample the number of type 1 or duplicate cases of type 2 to get equal values of each label. Balancing the labels prevent to create a bias machine learning model.
  - We also have to check if there is any missing data in any of our features column and replace them or either delete the corresponding row (in case of low number of Nan)
  - Moreover, most of the features have categorical values that need to be transformed to numerical values in order to improve the predictability of our model.
  - Finally, we will be using several types of supervised learning models and compare their performance to choose the best classifier for this project.

### 3 References

1. World Health Organization (WHO). Global Status Report on Road Safety 2018. December 2018.[cited 2019 April 8].  
[https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2018/en/external  
icon](https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/external_icon)