Capstone project: Car accident severity

By: Manu Sharma
Date: 11 October 2020

1. INTRODUCTION

1.1 BACKGROUND

According to a survey, in 2013 alone 54 million people worldwide sustained injuries from traffic collision. This resulted in 1.4 million deaths in 2013, up from 1.1 million deaths in 1990. About 68000 of these occurred in children less than five years old. Almost all high-Income countries have increasing death rates due to traffic collision. Middle- income countries have the highest rate with 20 deaths per 100000 habitants, accounting for 80% of all road fatalities with 52% of all vehicles. While the deaths rate in Africa is the highest i.e. 24.1 per 100000 inhabitants, the lowest rate is to be found in Europe i.e. 10.3 per 100000 inhabitants.

When driving, there are countless factors and hazards that can stand between you and your destination. Whether its torrential rains, bumper-to-bumper traffic that's at a complete stop and in worse case its wreckage. Given the opportunity, many drivers would like to have something that could give them warning to road conditions and their relation to the possibility of getting into a car accident, how detrimental it would be. This would allow drivers to adjust their route, drive more cautiously or even completely change travel plans.

1.2 BUSINESS PROBLEM

The world as a whole suffers due to car accidents, including the USA. National Highway Traffic Safety Administration of the USA suggests that the economical and societal harm from car accidents can cost up to $871 billion in a single year. According to 2017 WSDOT data, a car accident occurs every 4 minutes and a person dies due to a car crash every 20 hours in the state of Washington while Fatal crashes went from 508 in 2016 to 525 in 2017, resulting in the death of 555 people. The project aims to predict how severity of accidents can be reduced.

1.3 Stakeholders

Given the opportunity, many drivers would like to have something that could give them warning to road conditions and their relation to the possibility of getting into a car accident, how detrimental it would be. This would allow drivers to adjust their route, drive more cautiously or even completely change travel plans.

2. DATA UNDERSTANDING

There are a lot of problems with the data set keeping in mind that this is a machine learning project which uses classification to predict a categorical variable. The dataset has total observations of 194673 with variation in number of observations for every feature. First of all, the total dataset was high variation in the lengths of almost every column of the dataset. The dataset had a lot of empty columns which could have been beneficial had the data been present there. These columns included pedestrian granted way or not, segment lane key, cross walk key and hit parked car.

The models aim was to predict the severity of an accident, considering that, the variable of Severity Code was in the form of 1 (Property Damage Only) and 2 (Injury Collision) which were encoded to the form of 0 (Property Damage Only) and 1 (Injury Collision). Furthermore, the Y was given value of 1 whereas N and no value was given 0 for the variables Inattention, Speeding and under the influence. For lighting condition, Light was given 0 along with Medium as 1 and Dark as 2. For Road Condition, Dry was assigned 0, Mushy was assigned 1 and Wet was given 2. As for Weather Condition, 0 is Clear, Overcast is 1, Windy is 2 and Rain and Snow was given 3. 0 was assigned to the element of each variable which can be the least probable cause of severe accident whereas a high number represented adverse condition which can lead to a higher accident severity. Whereas, there were unique values for every variable which were either 'Other' or 'Unknown', deleting those rows entirely would have led to a lot of loss of data which is not preferred.

3. FEATURE SELECTION

- INATTENTIONIND: Whether or not the driver was inattentive (Y/N)
- UNDERINFL: Whether or not the driver was under the influence (Y/N)
- WEATHER: Weather condition during time of collision (Overcast/Rain/Clear)
- ROADCOND: Road condition during the collision (Wet/Dry..)
- LIGHTCOND: Light conditions during the collision (Lights On/Dark with light on)
- SPEEDING: Whether the car was above the speed limit at the time of collision (Y/N)