

Car Accident Severity Project

1. Introduction

According to the WHO¹ (World Health Organization), road injuries are the 10th cause of death worldwide. This becomes even more concerning when we realize it is the number one, non-disease cause of death in the world. Also, according to the U.S. Department of Transportation's Federal Highway Administration, weather and road conditions have a significant impact in the amount and severity of road accidents².

The big question is: How can we use data to prevent some car accidents and yearly deaths that come from it?

One possible application can be the use of Machine Learning to create a model that can, with a high degree of accuracy, predict the possible severity of an accident according to the weather and road conditions at a particular location and at a particular time. This could be used to forewarn drivers so that we have less vehicles on the road hence reducing the possibility of accidents and deaths.

Apart from public organizations like U.S. Department of Transportation, private companies might also be interested in such a predictive model, that could be integrated into already existing technology. For example, Google might put it to use together with their Google Assistant on Android Devices, to warn drivers as they are beginning their drive on using their scheduled calendar events, to stay home and reschedule by letting them know that current conditions are conducive of severe accidents, which can lead to heavy traffic in the best of cases to severe injury and death in the worst scenarios.

The main objective is to save lives and prevent severe car accidents through a predictive machine learning model that takes weather and road information into account and predicts the severity of potential accidents.

2. Data

The data set to be used for training is from the Collision Records from the Seattle Department of Transportation³, containing a sample of 194673 car crash events from 2007 to 2016, including the severity of the accident as well as many specifics of the weather and road conditions at the moment of the accident and other details like speeding, hitting a parked car and others.

¹ <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>

² https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm

³ <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

We will focus on Severity of the accidents as the target variable to predict, and for independent variables, we will focus on information relating to the weather and road conditions. We will leave out behavioral information like if the car was speeding or if the obstacle hit was parked, since these are not relevant to the predictive scope of the project and its use (getting people to avoid the road during conditions that lead to severe accidents).

All the independent variables related to weather and road conditions will be analyzed to determine the relation (positive or negative) to the target outcome (severity of the accident), and from there it will be determined which of them will be used. Then different algorithms will be applied including, but not limited to, Multiple Linear Regression, Logistic Regression and Clustering, and then evaluated to find the best fitting model that can most accurately predict the severity of an accident from weather and road conditions.