Predicting Car Accident Severity

Karan Y Thakur

October 2020

Introduction

Everybody enjoys going for a ride to a nearby hill station, a beach or visit a relative/friend in another city on a vacation or uses a road transport to attend a scheduled meeting in a different city. And often we see a horrific accident on the way and may get stuck in a terrible traffic jam. We also see the police, the ambulance or a helicopter transporting the ones involved in the crash to the hospital.

As per WHO, Road traffic crashes result in the deaths of approximately 1.35 million people around the world each year and leave between 20 and 50 million people with non-fatal injuries. More than half of all road traffic deaths and injuries involve vulnerable road users, such as pedestrians, cyclists and motorcyclists and their passengers. In addition to the human suffering caused by road traffic injuries, they also incur a heavy economic burden on victims and their families, both through treatment costs for the injured and through loss of productivity of those killed or disabled. More broadly, road traffic injuries have a serious impact on national economies, costing countries 3% of their annual gross domestic product.

The cause for an accident could be due to Environmental conditions like rain or a land slide or the reckless driving or a dangerous turn on the road or the road intersections. Nobody would want to be a victim of an accident. Wouldn't it be great if we have way or an application that alerts us to be cautious over a patch of the route or reschedule a planned trip. This could also help the government to enforce more safety protocols like placing the signals on the roads or speed breakers before the start of such zones. In order to achieve this, we would need to analyse the factors related to an accident and its severity.

The machine learning model could be built and integrated with the vehicle to intimate the driver in real-time before approaching the accident-prone zones.

Data

For the analysis, I will be using the extensive data set provided by the Seattle Police Department, which is a labelled data with over 190,000 observations collected since 2004. This data can be used to understand factors related to the accidents severity by building a statistical model such as a supervised machine learning classifier.

The dataset has details related to the accident such as:

1. The Collision type, the total number of person, pedestrian, bicycles, vehicles involved in the collision, whether the driver was speeding or not.
2. The location details of the collision: latitude and longitude coordinates, location type, the type of junction.
3. The environmental features like: Weather, Road Condition, Light Condition
4. The state of driver if he was attentive or not, was he under the influence of alcohol, whether he was speeding or not.