IBM

DATA SCIENCE CAPSTONE

PRELIMINARY REPORT

Introduction

Road accident fatalities are one of the major causes of death and injuries across the world. Every year more people die due to road accidents than they do to diseases. According to WHO Reports, approximately 1.35 million people die due to road traffic accidents across the world. Not only are road accidents resulting in loss of life, it is also hurting the economy; as traffic accidents cost most countries around 3% of their GDP. Moreover, more than half of all road deaths are among vulnerable road user; like bikers, pedestrians and cyclists.

Accidents may not be entirely human fault, as many factors culminate to cause a traffic accident, for instance a foggy weather with reduced visibility range, a road wet after a rainy day, etc.

So, what if we can predict (with certain accuracy) the same and prevent this from happening, or suppose the inevitable has happened, and we are able to send a signal immediately to the nearest hospital about this incident for immediate response; or what if you are on your way to your office and have a way to prevent getting stuck in a traffic jam due to a road accident.

With this as my background, I would like to proceed with this report.

Problem Statement

So, this project is driven by the following underlying idea:

1. How do we create an algorithm using a data-driven approach that can predict occurrence of an event (here it is a traffic) with an acceptable accuracy.
2. How do we make it so that it can be available to the general public as well as to the concerned authorities.

The second part of the problem statement will require some degree of explanation, but for the first part, we can utilize Machine Learning models.

This is a preliminary report where I am supposed to provide you with the background of my project. The accompanying code (that is uploaded on my GitHub <https://github.com/LordNex0831/Codes.git>) will provide you with some preliminary analysis and wrangling operations that I have preformed on the dataset used for this project (link for the dataset is *https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv*).