Predicting Car Accident Severity

Applied Data Science Capstone Project
IBM Data Science Professional Certificate

Mustafa Bal

1. Introduction

1.1 Background

As one of the leading international institutions that has set road traffic accidents as one of its primary agendas on global scale, the World Health Organization (WHO) reports that roughly 1,35 million people die every year as result of road traffic accidents (WHO, 2020). Road traffic accidents also leave between 20 - 50 million people injured, some of whom suffer long term health problems and some physical disability or impairment. According to the Center for Disease Control and Prevention, In the United States, where approximately 10 million individuals are involved in a car accident annually (Rolison et al., 2018), car accidents are the number one cause of death for teenagers. Unfortunately, the numbers are not different for other countries. According to the records of the WHO, injuries that inflicted as a result of traffic accidents are the leading cause of death for children and young adults between the ages of 5 and 29.

Road traffic accidents result in also significant economic losses for states and for the individuals and their families who are involved in accidents. The WHO estimates that the cost of road traffic accidents to most countries to be approximating 3% of their gross domestic product. We cannot also ignore the long-term social consequences of traffic accidents on especially children and families. In brief, as a serious global problem road traffic accident should be on the agenda of researchers.

1.2 Problem Statement

As a serious global problem road traffic accident has been on the agenda of researchers as well as governments. I believe to tackle with this issue effectively we should develop a coordinated and multi-disciplinary approach. In this respect, this study would like to contribute to the analysis of the road traffic accident problem by employing tools and methodologies of Data Science field.

This study aims to develop a model that correctly identifies the severity of the traffic accidents in the city of Seattle. To this end, a relatively rich road accident records dataset was analyzed by using several different machine learning models.

1.3 Interest

Developing a machine learning model that predicts the severity of road traffic accidents with a relatively high rate of accuracy would be certainly beneficial for everyone. Moreover, since machine learning models, with complex algorithms enable researchers to identify salient factors that contribute to accidents and their severity. Such factors are sometimes not easily discernable among large datasets.