Predicting the Severity of Car Accidents

Youssef Saleh

September 3, 2020

1. Introduction

1.1 Background

Car accidents without doubt are one of the most tragic incidents that take place repeatedly, they lead to losses in both money and lives. Although many governments have tried hard to prevent accidents from taking place through safety precautions and organizational matters, accidents still take place and results in heavy damages and costly losses. Moreover, most of car companies have developed safety measurements and installed new technologies into the vehicles, which has successfully in recent years reduced the number of casualties during car accidents and also sometimes helped in preventing accidents from taking place in the first place. However, it will be also useful if governments were able to group accidents according to their severity by taking into consideration various variables, this will enable governments to understand the reasons for each accidents group and concentrate the efforts on dealing with these reasons and specially with the most severe accident group. Then it would be beneficial to predict the severity of accidents in the future and try to deal with causes.

1.2 Problem

Dataset that will help in determining the severity of the car accidents will contain various attributes representing many aspects and conditions of car accidents like: weather, road, driver's condition, etc. All of this will help in the aim of this project which is predicting the severity group of car accidents.

1.3 Interests

Obviously, this project and the created model would be very helpful for governments, as they can use it as reference when establishing new roads, reorganizing traffic of the cities or trying to understand the main reasons behind different severity of car accidents to deal with them. Although attributes of accidents differ from country to another, this project illustrates to any country, how to implement and create a model with their own conditions.

2. Data Acquisition and Data Cleaning

2.1 Data

There are many sources to get such dataset, however a sample dataset provided by Corsera will be used during the course of this project. This dataset has 38 attributes and contains column for accidents severity (our target). Moreover, the significance of attributes will be assessed and then chosen accordingly. The unused attributes will be dropped out of our dataset and the missing values will be dealt with during preprocessing. Since grouping of already labeled accidents severity is the objective, Classification learning algorithm will be implemented (Supervised learning) to create the intended model.