CAR ACCIDENT SEVERITY

Capstone project

Week 1

1. Introduction / Business problem

1.1. Record

Road Traffic Injuries (RTIs) are a major public health problem. The World Health Organization (WHO) reports that the number of deaths due to road accidents has exceeded one million in recent years. (1)

The problem is huge and concerns everyone, a fact that makes the analysis of road accidents necessary. Forecasts have been rising in recent decades. (2) The analytical approach to the data will provide answers that can be used to identify and predict factors that affect the severity of road accidents.

The most popular scientific method is Machine Learning, because it has the ability to identify existing patterns in data and to predict, through the creation and evaluation of different algorithms.

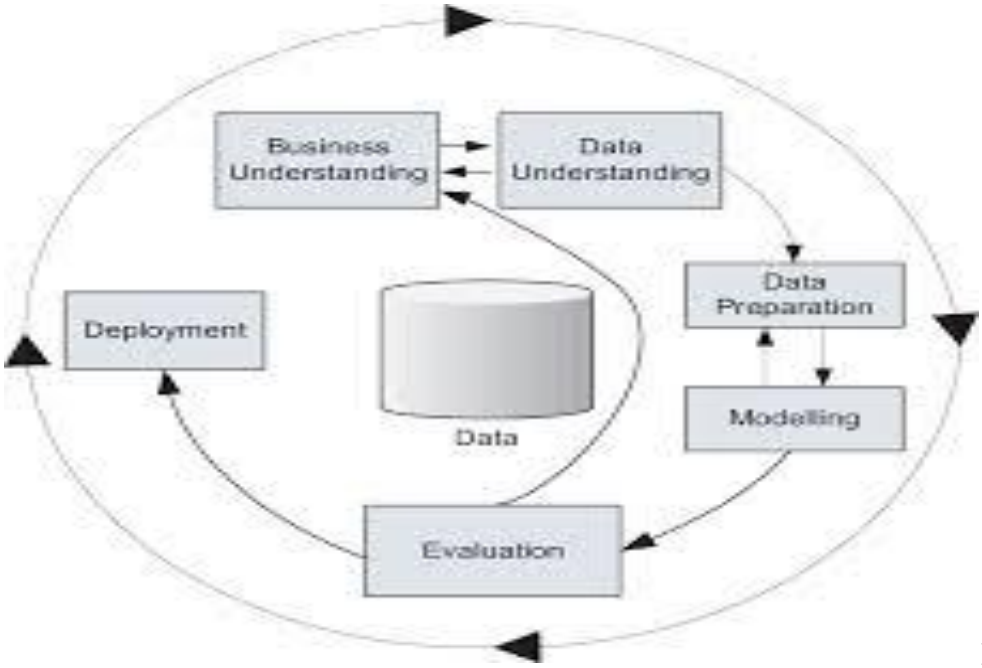
Furthermore, Machine Learning can manage another problem that arises, the large amounts of data that are generated, since road accidents are frequent and increasing at a rapid rate.

1.2. Purpose

The aim is therefore to predict the severity of the accident through training and assessment machine learning algorithms, with the help of a data set, including recording provided by SDOT Traffic Management Division, Traffic Records Group in Seattle, United States.

As an approach to achieve this goal, the interprofessional standardized data mining process (CRISP-DM) (Figure 1) will be

applied. The data will therefore be well understood and prepared before being fed for the forecast modeling analysis in the next steps.



F1, (3)

1.3. Interested

The analysis is addressed to those involved in road traffic such as:

- National Emergency Center,
- Hospitals,
- Road managers,
- Traffic control authorities,
- Roadside assistance services

in order to guide them to improve safety margins and ways to deal with the severity of the accident.

References:

(1) World Health Organization: http://www.who.int

(2) Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights, S. Moosavi et al., SIGSPATIAL '19, November 5–8, 2019, Chicago, IL, USA

(3) Cross-industry Standard Process for Data Mining (CRISP-DM), https://en.wikipedia.org