Capstone Project

REPORT: Predicting Car Collision Severity in Seattle

By Ramy Moussa
August 2020

1. Introduction: Business Understanding

1.1 Background

For any traffic participant, pedestrian, cyclist, or motorist, an accident is an unexpected and undesired thing to experience. Property damage, personal injury or death as consequences of collisions not only impact the lives of individuals but also society and economy at large. When cars were first introduced in the US at the beginning of the 20th century, they were few numbers but the fatalities they caused were many. Neither comprehensive traffic laws nor significant safety features in cars or on roads existed. Since then, landmarks in auto safety (produced by commercial, technological, legal, or moral forces) have reduced both motor vehicles accident deaths and pedestrian deaths to an all-time low (see Figure 1 and Figure 2).[1] To name a few such landmarks: car manufacturers have introduced seatbelts and airbags; civil engineers have paved roads and added reflecting guard-rails to highways; moral campaigners have lobbied against drunk driving; and last but not least government has promulgated and enforced many traffic (safety) laws. For comparison, in 2015 almost 5,000 pedestrians died in traffic accidents, whereas in 1937 15,000 pedestrians were killed, when the US had far fewer cars and two-fifths of its current population (see Figure 2). Although motor vehicle accident deaths and pedestrian deaths in the US are at an all-time low, each death or injury that still occurs is a tragedy and thus remain a public health concern.[2]

Figure 1: Motor Vehicle Accident Deaths, US, 1921-2015

Source: Pinker, Steven. *Enlightenment now*. Penguin, 2018.
http://www.informedforlife.org/demos/FCKeditor/UserFiles/File/TRAFFICFATALITIES(1899-2005).pdf.
http://www-fars.nhtsa.dot.gov/Main/index.aspx.
https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812384.

[1] Both figures are taken from Pinker, Steven. *Enlightenment now*. Penguin, 2018, p. 222-225.
[2] The WHO defines it as public health issue that causes approximately 1.35 million deaths around the world each year and leave between 20 and 50 million people with non-fatal injuries. https://www.who.int/health-topics/road-safety.

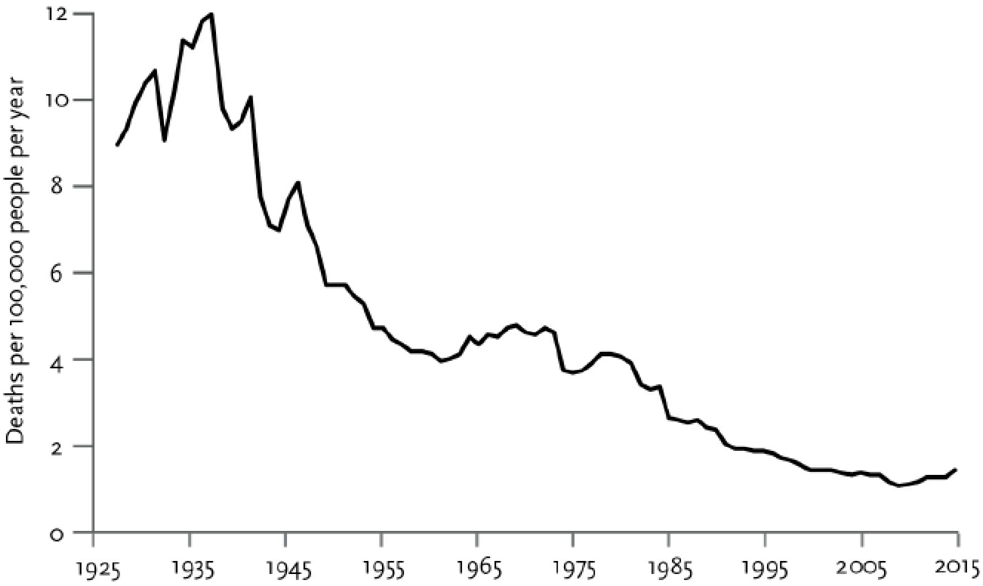

Figure 2: Pedestrian Deaths, US, 1927-2015

Sources: Pinker, Steven. *Enlightenment now*. Penguin, 2018.
For 1927–1984: Federal Highway Administration 2003.
For 1985–1995: National Center for Statistics and Analysis 1995.
For 1995–2005: National Center for Statistics and Analysis 2006.
For 2005–2014: National Center for Statistics and Analysis 2016.
For 2015: National Center for Statistics and Analysis 2017.

1.2 The Objective

Before AI-driven cars will become ubiquitous and hopefully reduce traffic accidents to near zero, this report argues that Machine Learning Algorithms are the next important step to reduce traffic accidents. Machine Learning Algorithms can analyze historical data on traffic collisions and determine what features can best predict the occurrence and severity of accidents. The immediate application is that traffic authorities can use the algorithm's output in combination with electronic traffic signs to warn motorists (and other traffic participants) about potentially dangerous traffic conditions. The future application is that such an algorithm can update AI-driven cars on traffic condition in the area (in addition to the data that the car collects and analyzes itself) to improve its driving and the safety for all traffic participants.

As a case study, this report will use the Collision Data collected by SDOT Traffic Management Division, Traffic Records Group (from 2004 to present) for the city of Seattle to build a classification model (a supervised machine learning algorithm) for predicting the severity of collisions.[3] Following best practice and to ensure reproducible results, this report uses the Cross-Industry Standard for Data Mining methodology (CRISP-DM, see Figure 3).[4] In the second chapter, this report will investigate the Collision Dataset to determine the potential features for a machine learning model by performing an exploratory analysis on the attributes. The third chapter uses the insights from the second to pre-process all relevant feature variables and the target variable putting them into final dataset ready for machine

[3] See https://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0.
[4] See https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining.

learning. Chapters four and five will serve to build various type of classification machine learning models and to evaluate their performance in predicting collision severity. Given the limitations of this report, the final chapter will discussion deployment options for potential stakeholders (i.e. public traffic authorities, emergency services, or car manufacturers).

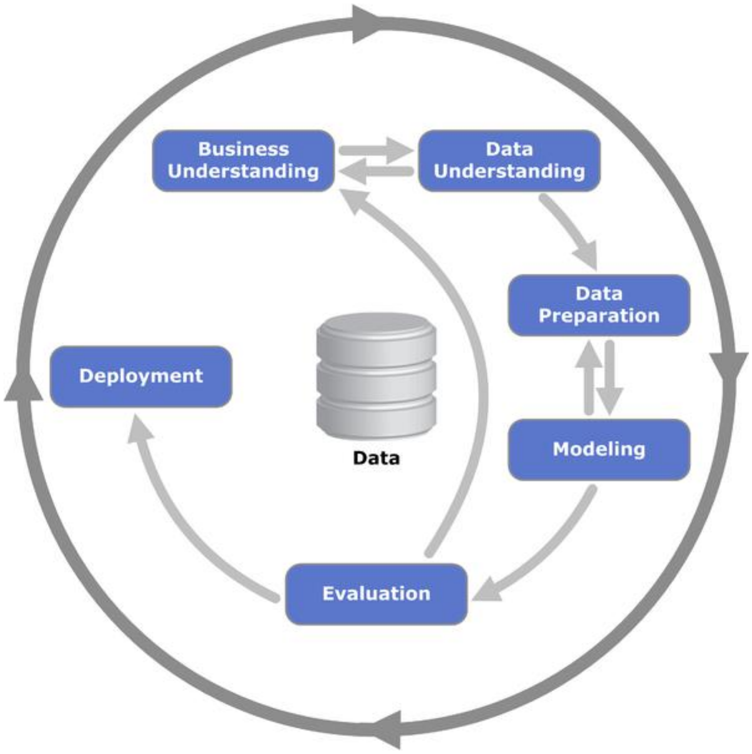

Figure 3: Cross-industry standard process for data mining, known as CRISP-DM

Sources: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

The following chapters will appear in the full report:

2. Data Understanding

3. Data Preparation

4. Modelling

5. Evaluation

6. Deployment