

A.I. & Data Science Challenge 1: Red life: Predicting Casualty Severity in Traffic Accidents



University of Applied Sciences

Delivery Document

Mateusz Mierzejek

2775409

12/20/2021

Introduction

Every part of a data science project needs to conclude. And hopefully with beneficial results. Though a model might not be the most appropriate way to display results to a client. Sometimes it can give certain insights into situations.

On the other hand, a model can be a great way of implementing a real-world use case for an algorithm. There are various frameworks nowadays for implementing these results.

These things can be possible with a successful project. The next step towards this is interpreting the data and finding future solutions.

Contents

Introduction	2
Collection & Evaluation	4
Final Impact Assessment	5
Deployment Recommendation.....	6
Concluding Chapter	7

Collection & Evaluation

In the proposal document we started out by exploring different possibilities and factors which might have an influence in the severity of a traffic incident. Transiting from a general view of Europe and focusing on the Netherlands. We found this data not to be of sufficient quality based on the factors of significant amount of unknown or missing data.

By analyzing these datasets some valuable insights and techniques have come to play. Like the selection of features and data visualization. In order to have a proof of concept, the choice has been made to opt in for a British dataset. This had the benefit of little to no missing data. And thus, a bigger selection of features which could be considered important when predicting the severity. Some of the main features being:

feature	importance
age_of_casualty	0.068232
age_of_driver	0.054631
engine_capacity_cc	0.047861
generic_make_model	0.043555
age_of_vehicle	0.041958
casualty_imd_decile	0.039612
vehicle_direction_to	0.038464
driver_imd_decile	0.037849
vehicle_direction_from	0.037550
speed_limit	0.034799
vehicle_manoeuvre	0.032502

We observe the **age** of the *casualty* and *driver* have a high impact. This can be because a combination of some age groups leads to more dangerous scenarios.

Another important factor is the data about the vehicle. We can see the **engine capacity** and **model** play a part in the severity of an incident.

In addition to that an interesting benefactor are the **imd_decile** features. These describe the relative deprivation in the living area of the individual

Finally, there are some descriptive features like the **direction** the vehicles was heading and some limitations like **speed limit**.



Use has been made of multiple algorithms, but Random Forest Classifier scored the best. A reasoning for the high accuracy is over sampling of the dataset. This might have had an influence on the model being most biased.

We observe the train score being higher than the validation score. This is a sign of overfitting. Thus, another approach would have to be taken to ensure the data fits the requirements.

Final Impact Assessment

The data that has been gathered in this research has given insight into possible scenarios where the casualty severity is high. Based on the selected features potential measurements can be done within the existing traffic infrastructure. Benefitting in both less incidents and lower casualties.

The model itself can be considered as a tool for gaining insight into current situation within traffic and how to avoid them. The data could be use for targeting specific participants in traffic for observation and/or containment of danger.

Deployment Recommendation

The deployment of the model would see no further added benefit to the research that has been done. With the progress of this project, it became clearer that it is not applicable in a real-world scenario, but rather an indication or clarification for the points of interest within traffic.

Tracking these features could lead to a better understanding of the nature of incidents and how they can be prevented.

Concluding Chapter

To finalize the documents provided will give further insight into the research that has been done in the order of:

1. **Project Proposal.pdf** along with **Traffic Casualty Severity EDA.html**
Evaluation of project and domain understanding along with EDA
2. **Traffic Incident Casualty Severity Phase 2 + 3.html**
Data preparation, Data visualization, Modeling
3. **Delivery.pdf**
Evaluation