Exploring Vehicle Accident Severity Levels

A description of the problem and background:

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product.

In high income countries, the combination of safety standards and education have made tremendous strides in automobile safety. The challenge now is to ensure the low and middle income countries which now account for 50% of new car sales and production and over 90% of road fatalities can be brought up to global minimum standards.

As much knowledge as we have relating to vehicle safety, most still struggle to ensure they are doing everything necessary to guarantee their drivers are safe on the roadways and limit their exposure to crashes.

A description of the data:

The data set which will be used is the “Collisions—All Years” provided through the IBM Data Science Capstone by Coursera. They are data from SDOT Traffic Management Division.

The data consists of the information about the accidents. This includes all types of collisions from 2004 to present. The data has 194,673 rows and 38 columns. The columns are referred as the features of the dataset and every row referred in a single collision.

METHODOLOGY

The business problem:

Road traffic injuries can be prevented. Governments need to take action to address road safety in a holistic manner. This requires involvement from multiple sectors such as transport, police, health, education, and actions that address the safety of roads, vehicles, and road users. Effective interventions can be done if we make good use of billions of collected data and the power of data science, building models which predict the severity of vehicles’ accident.

Data understanding and preparation:

In this phase, we used the “Collisions—All Years” data set and we gave our attention to above mentioned features in order to explore their correlation with the severity of the collision. We discover that there are some data that was described as “unknown” or “other” using simple functions and pandas for cleaning dataset.

Data analysis:

A screenshot of a cell phone

Description automatically generated

These two figures show most car accident happened in clear and dry weather, which may be different from our common senses.

A picture containing stationary, pencil

Description automatically generated

These figures show the number of car accidents decrease by year, in 2020 it’s the fewest that’s because it only include data by now. in month of car accidents happened, it seemed average, in day the car accidents happened, the beginning of the month seems have more accidents, in hour of car accidents happened, afternoon has the most car accidents.

Modeling:

First we needed to convert the categorical columns of the selected features into numeric values so we can process them in our models. In this phase, various algorithms and methods was selected and applied to build our model. We chose a rate of 70-30 on train and test our models.

The selected methods was :

Logistic Regression

Decision Tree

Random Forest Classifier.

Evaluation:

The model needs to be evaluated thoroughly to ensure that the business objectives are achieved. Certain metrics used for the model evaluation.

We chose:

Accuracy

Recall

F1-score

Precision

Results:

• There are no significant correlation between the attributes and the severity of collision.

• The results of the different Machine Learning methods are the following:

A screenshot of a cell phone

Description automatically generated

Discussion and Conclusion:

In our minds, there is tight correlation between weather conditions, road conditions, light conditions, junction types and severity of vehicles collisions. In reality there are so many attributes that affect the outcome of a crash so that the contribution of a specific one is too hard to be evaluated.

It remains the human influence the most significant factor for the result of an accident. The level of watchfulness, the driving experience and even the driving habits are difficult to be measured and to be part of a model.