Accident Severity Predictor – Seattle Traffic

APPLIED DATA SCIENCE – CAPSTONE IBM/COURSERA

FUTURA DATASCIENCE CORPORATION – PRAKASH SHETTIGAR

Business Problem

In the United States and throughout much of the world, car accidents are a leading cause of serious injury and death. In fact, in the U.S. alone, at least 38,800 people were killed in motor vehicle collisions in 2019.

In this project we will try to determine or predict the severity of the traffic accident. Specifically, this solution or analysis will be targeted to the following stakeholders.

A. Mobile Map applications

B. Vehicle insurance providers

C. Department of Motor Vehicles and other government bodies

Data Exploration

Based on the definition of our problem, factors that will influence our decision are:

- driving under the influence of alcohol or other substances
- speed of the vehicle
- weather / light conditions
- road conditions

Methodology

Split the data set into and training and test sets. Used 70/30 ratio for data split.

We will apply the following **machine learning** algorithms on the training set:

- **K-means Clustering**
- **Decision Tree**
- **Logistic Regression**

And evaluate the test set on the trained model. Verify the accuracy with Jaccard index and F1 score.

Results and Discussion

Our analysis evaluated various machine learning models such as **K-means Clustering**, **Decision Tree**, **Logistic Regression** to determine if accident severity can be predicted based on the factors such as **month**, **road conditions**, **weather conditions** and **light conditions**.

Jaccard index score was 67% - our test set and predicted result test matched to a reasonable degree. F1 score or **accuracy rate** of prediction was **54%**. Prediction accuracy using this analysis is not reasonable and may not be acceptable to the stakeholders.

Conclusion

Purpose of this project was to identify if there were certain factors which lead to more severe accidents. We used machine learning models to see if there is a pattern to predict the accident severity.

Our results were inconclusive and there were no clear indicators on whether certain factors lead to more severe accidents than others.

We will need to revisit the original dataset and research other attributes/parameters to see if these results could be improved upon. We might also need to pull other relevant datasets for more detailed analysis.