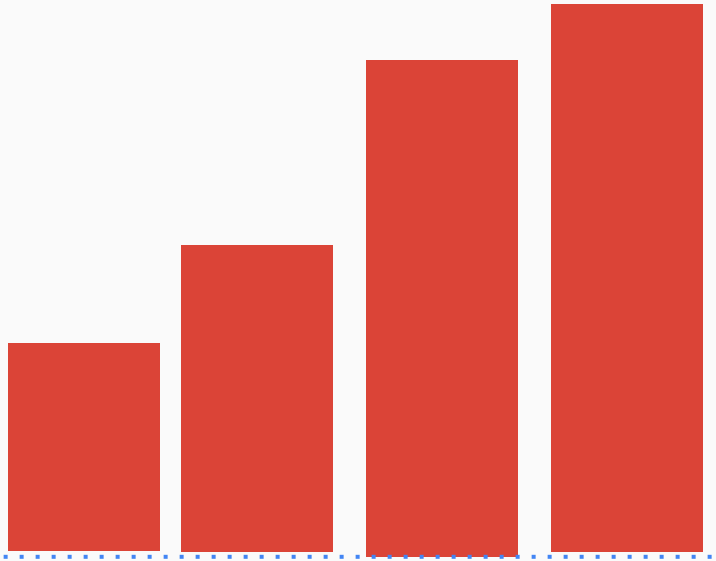
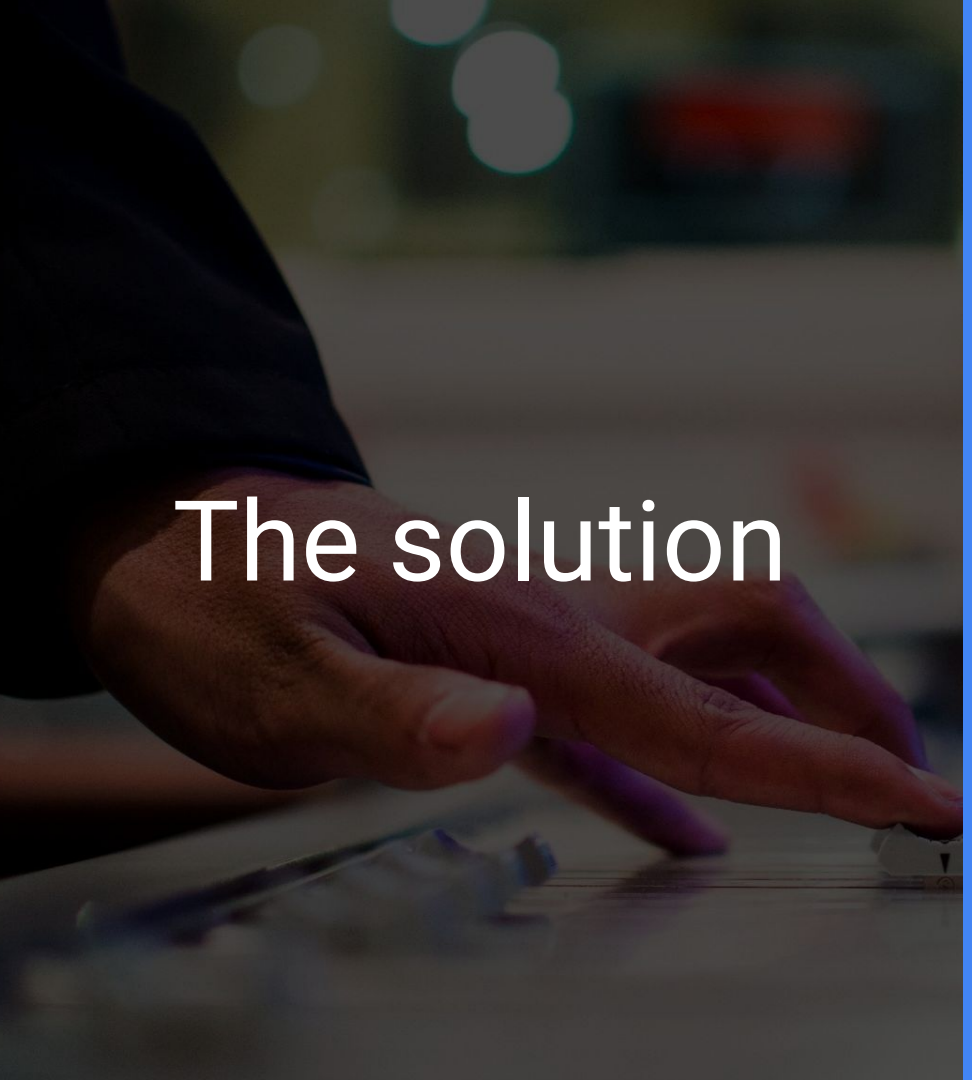
Car Accident Severity Analysis

The problem

The lives of approximately 1.35 million people are cut short as a result of a road traffic crash.

For children and young adults aged 5-29 years this is the leading cause of death.





The solution

Developing an algorithm to predict the severity of an accident given the current weather, road and visibility condition.

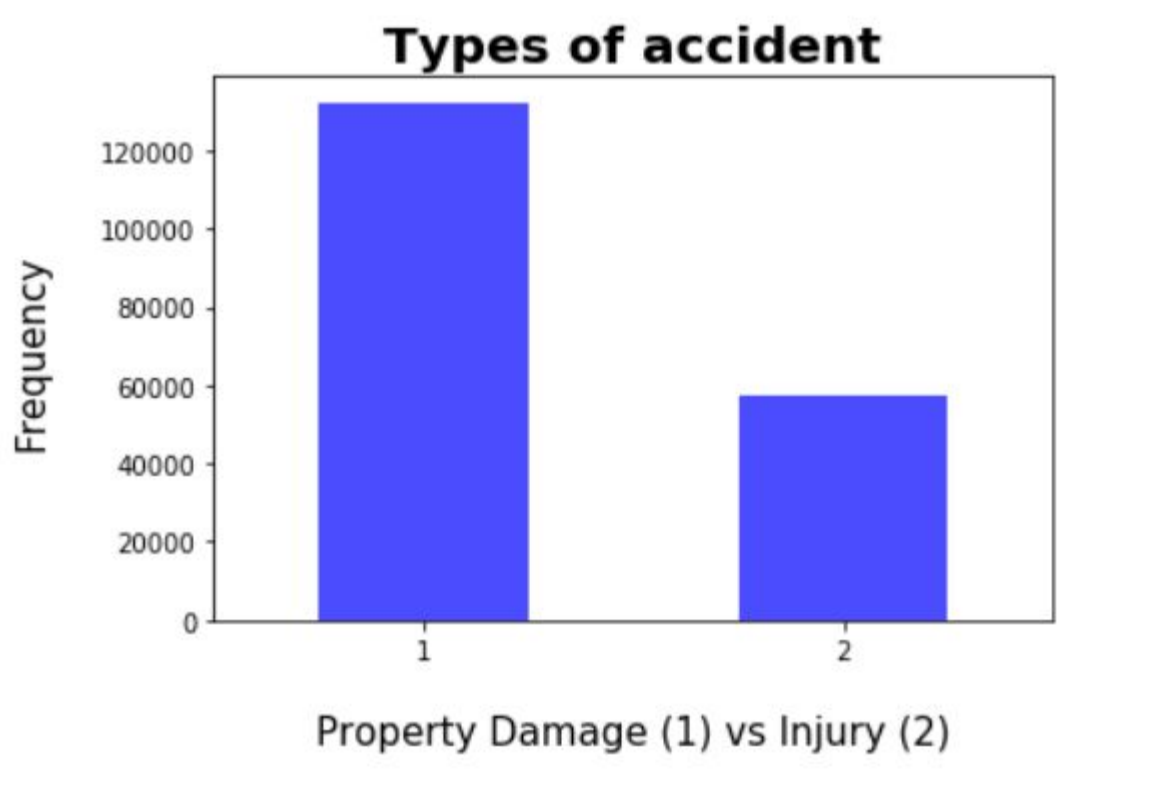
Data

The was data collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.

The variable severitycode, classifies the level of severity caused by an accident as:

1: Property Damage

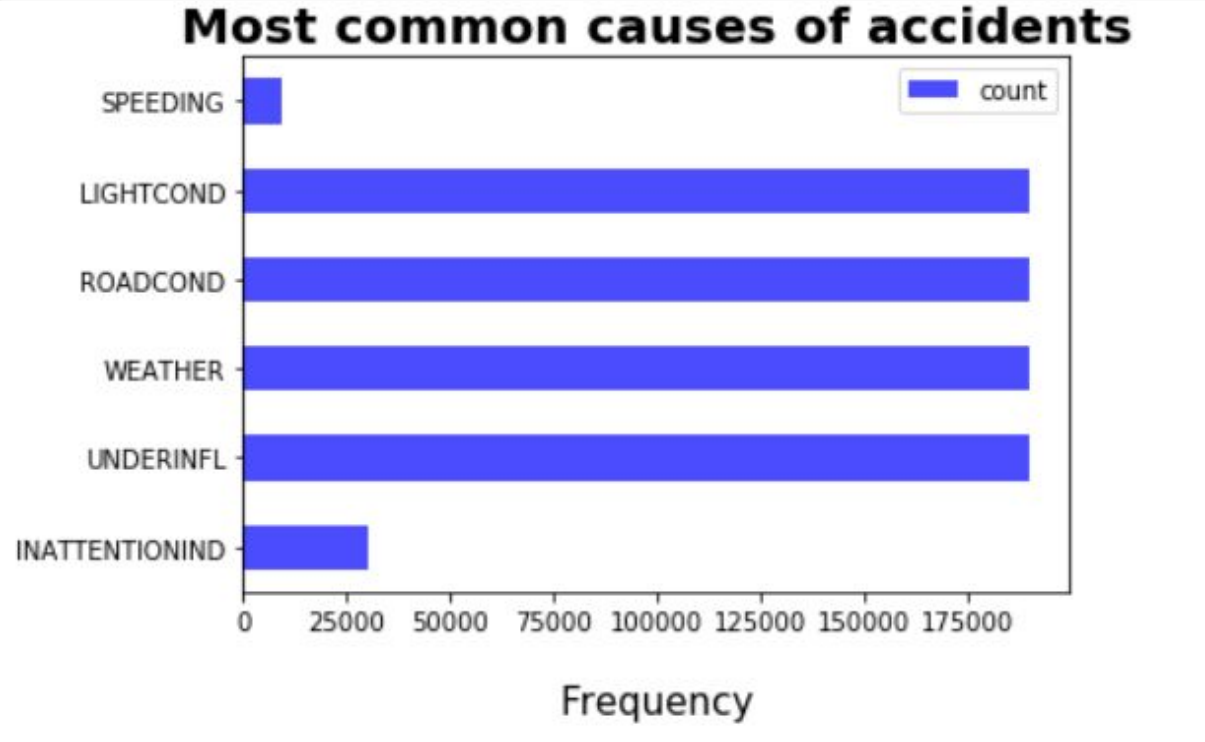
2: Chance of Injury



Data

The dataset provides a lot of different information about the circumstances in which the accidents took place.

Light Condition, Road Condition, Weather and Under Influence are the main causes of accident.



Methodology

Preparing the data

Select the important variables, clean Nan values and balance the dataset

Creating ML models

K-Nearest Neighbour (KNN), Decision Tree, Logistic Regression

Comparing

Compare the results and decide what model is the best

Training and testing

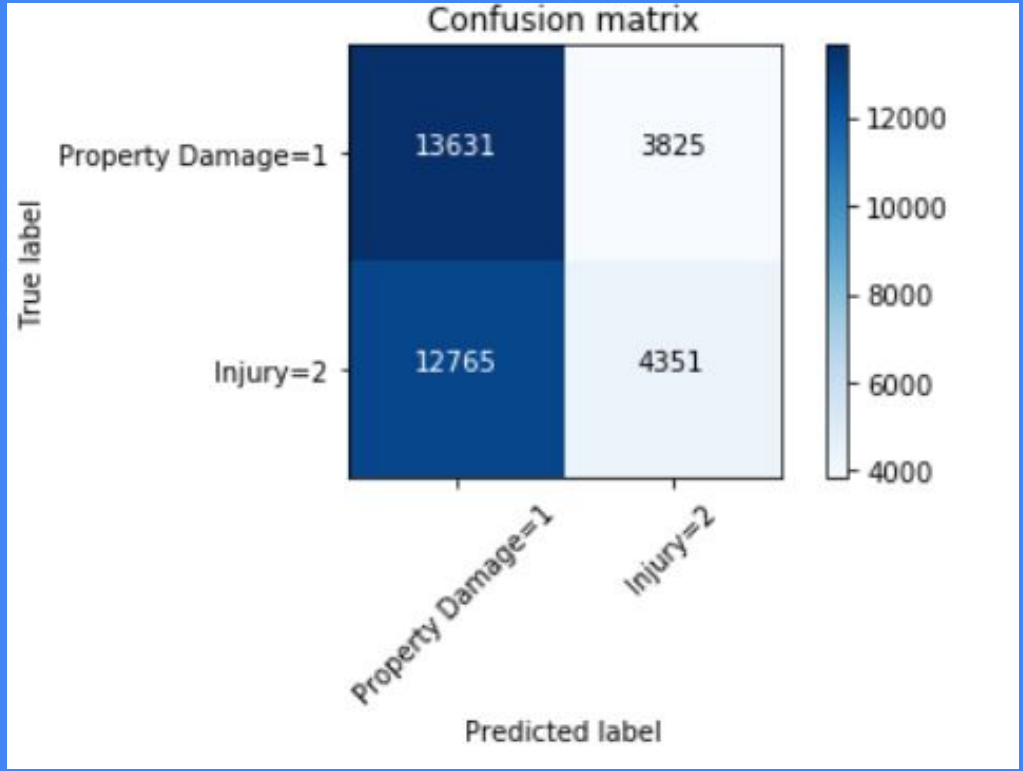
Split the dataset into one Train Set and one Test set

Evaluation

evaluate the results using Jaccard Score, T1 Score and Accuracy Score

Results

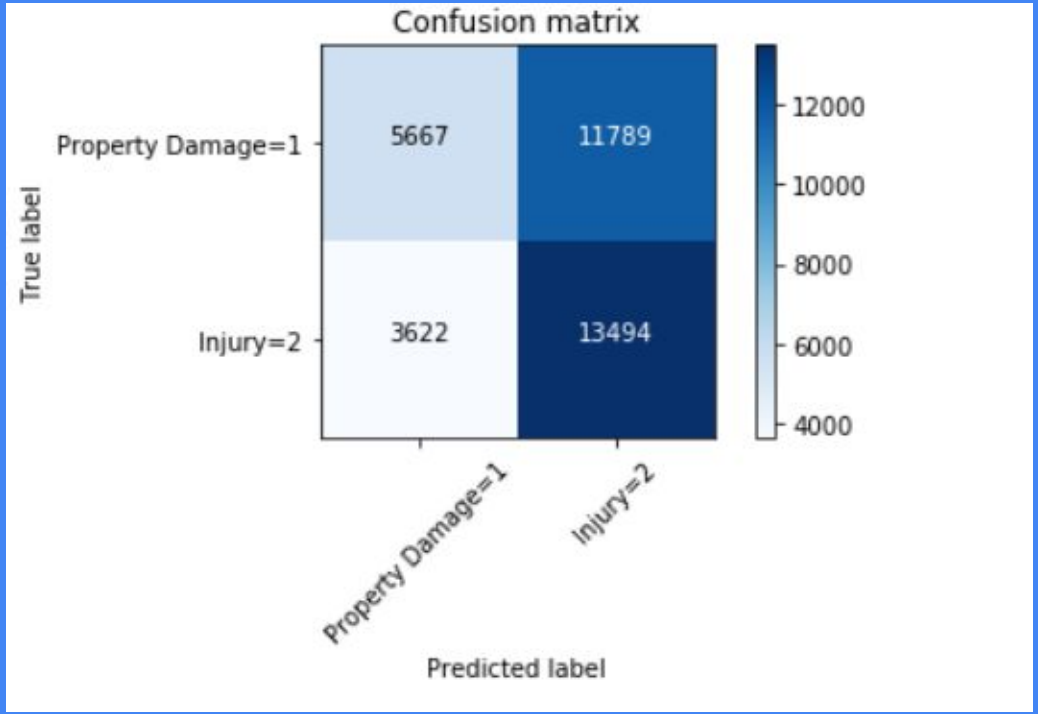
K-nearest Neighbor (KNN)



Jaccard Score	T1 Score	Accuracy Score
0.52	0.48	0.52

Results

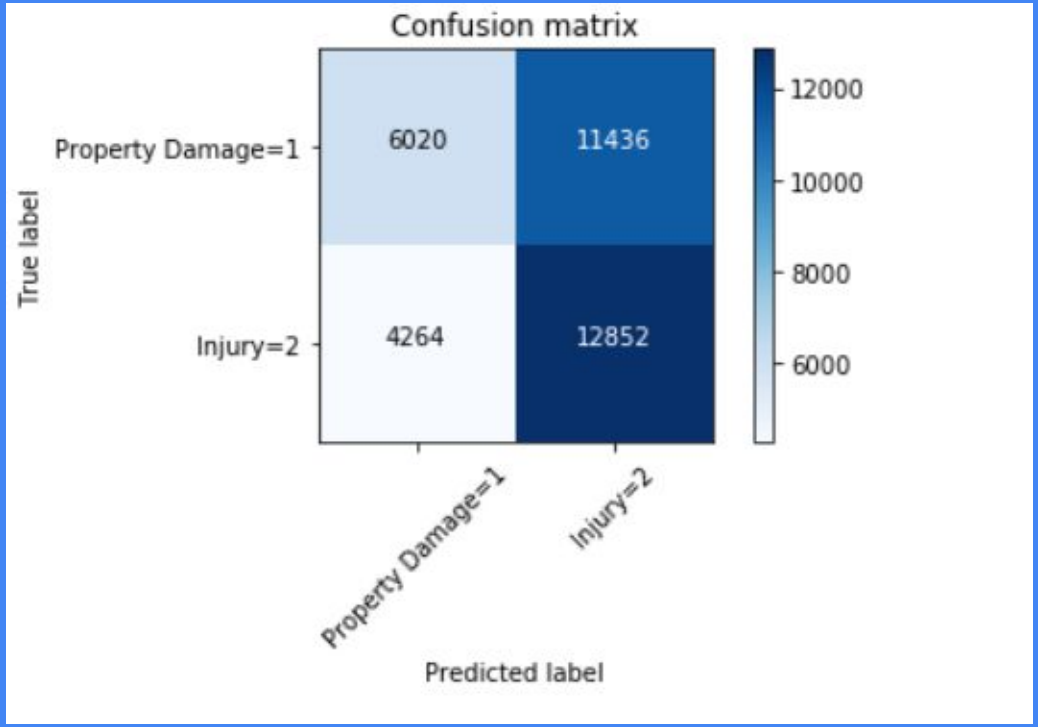
Decision Tree



Jaccard Score	T1 Score	Accuracy Score
0.55	0.53	0.55

Results

Logistic Regression



Jaccard Score	T1 Score	Accuracy Score
0.54	0.52	0.54

Conclusion

Jaccard Score	T1 Score	Accuracy Score
0.52	0.48	0.52
0.55	0.53	0.55
0.54	0.52	0.54

Multiple factors that can contribute to the severity of a Car accident.
After analysing the Jaccard Score, F1 Score and Accuracy score for the three Machine Learning models developed, it's noticeable that the Decision Tree is the best option.