



# Accident Severity Prediction

Classification Model



# *Importance of this Prediction*

These days vehicle accidents are getting very serious issue globally leading to deaths and serious injuries. Here we are concerned about car collisions and will try to understand the factors on which collisions happens or occurs.

Problem occurs with business understanding of the above mentioned issue. We need some solution to reduce or eventually stop this collisions. In an effort to reduce the frequency of car collisions in a community, an algorithm must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful.

# *Data Acquisition & Cleaning*

Our predictor or target variable will be 'SEVERITYCODE' because it is used to measure the severity of an accident from 0 to 1 within the dataset. Attributes used to weigh the severity of an accident are '**WEATHER**', '**ROADCOND**' and '**LIGHTCOND**' etc.

Severity codes are as follows:

0 : Little to no Probability (Clear Conditions)

1 : Very Low Probability - Chance or Property Damage

In its original form, this data is not fit for analysis. For one, there are many columns that we will not use for this model. Also, most of the features are of type object, where they should be numerical type.

We must use **label encoding** to convert the features to our desired data type. With the new columns, we can now use this data in our analysis and ML models.

We will use the following models:

### **K-Nearest Neighbour (KNN)**

KNN will help us predict the severity code of an outcome by finding the most similar to data point within k distance.

### **Decision Tree**

A decision tree model gives us a layout of all possible outcomes so we can fully analyze the consequences of a decision. In context, the decision tree observes all possible outcomes of different weather conditions.

### **Logistic Regression**

Because our dataset only provides us with two severity code outcomes, our model will only predict one of those two classes. This makes our data binary, which is perfect to use with logistic regression.

We have evaluate our models by following metrics:

1. For Decision Tree(Accuracy Score,F1 Score, Jaccardian Score)
2. For KNN (Accuracy Score,F1 Score, Jaccardian Score)
3. For Logistic Regression(Accuracy Score,F1 Score, Jaccardian Score & Log Loss)

# *Conclusion*

Based on historical data from several conditions pointing to certain classes, we can conclude that particular conditions have a somewhat impact weather or not on travel could result in property damage (class 0) or injury (class 1).