# Car Accident Severity Prediction

# **Introduction**:

# The purpose of the capstone project is to help people to travel in safe roads around their neighborhood based on the weather conditions. It will help people reach their destinations or jobs safer in desired time. This project will help people to know of their road conditions based on the present weather and how safe it is to take the routes. It will help the people who mostly rely on road trip and car travels for their daily jobs or travels to nearby city which they are not familiar. The project will warn traveler to avoid the situation of getting in to accident in uncertain weather conditions. It will help them to avoid unnecessary risks and relaxed travel.

# **Data:**

We will use the sample data set provided in the course. The data consists of Seattle city accident survey.

The data set contains the severity field which explains the fatality of an accident. The data set contains various fields that help to predict the severity label such as weather conditions, condition of a road, severity desc, light conditions, as well as collision type. Most of the data are categorical types and the data engineering process is needed to normalize the data to build the model.

Methodology:

The data has been analyzed with various parameters; like some of the columns have unknown values and it might affect the model fitness. The data has been plotted, against severity code and weather conditions, most of the accidents took place in wet conditions. Like said above, data is being plotted against multiple factors, road conditions, light conditions etc. The chosen test data has multiple factors to train the model and to predict the severity.

Model has been built based on classification algorithms. KNN, Decision tree, Support Vector Machine, Logistic Regression algorithms has been used, and the best performance algorithm is decided based on the jaccard score and F1 score.

Based on the metrics, all algorithms mostly returned with similar score, the data looks biased. The data has to be chosen with multiple variations to find the best model. So the data has been chosen with multiple limits and combinations, the model have been built on.

Results:

The metrics are being calculated and predicted results are being displayed in the notebook. Based on the metrics , Support vector machine model is best performing having the jaccard score of 1.

Observations:

Since the sample dataset is large, it’s better to split and build the model with different parameters.

Conclusion:

The car accident severity has been predicted using classification models.

# 