

# **INTRODUCTION**

Car accidents occur all over the world and cause several injuries and death in many cases.

Using data analytics, we can provide a solution that would help reduce magnitude of car crashes and even avoid further damage. We can predict accident severity using features such as weather, road, light conditions as well as number of people, vehicles involved in the car crash and other elements. We can then develop an application that allows drivers to be alerted of the severity level when the car crash is labeled severe (Based on label code).

In this case study, using the example of the Seattle Police Department, policemen, ambulance/rescue dispatchers, car insurance companies and drivers can gain lots of benefits using this application. They can increase their efficiency, effectiveness, reduce time, take more precise and fast decisions and of course protect the society.

## **DATA**

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present. The data consists of 38 independent variables and 194,673 rows. The target variable, "SEVERITYCODE", contains 2 categories (Injuries or Property Damage).

## **METHODOLOGY**

First, I acquired the data and applied the dataset in pandas. I then performed data exploration to understand the data, the most important features, what are the elements that are always present in car crashes with the highest frequency, the peak time of car crashes.

I then performed data cleaning, several features contained missing values, wrong formats, and categorical variables were transformed to encoded labels. I have then deleted the features that do not contribute to the model.

Finally, I applied different machine learning algorithms and evaluated the accuracy of each in the prediction as well as the features that have the highest predictive significance.

## **RESULTS & DISCUSSION**

After analyzing the data here are the findings I obtained:

- . Most accidents involve property damage rather than injuries
- . Most accidents occur on blocks rather than intersections
- . Pedestrians are rarely affected by car crashes
- . Most of the times, accidents involved 2 vehicles / 2 individuals
- . Highest number of accidents occur with parked cars
- . The weather does not have a high impact on accident frequency
- . Highest number of accidents occur when 1 car is moving the other parked, Entering at angle & rear end crashes.
- . From Logistic Regression, Decision Tree, KNN, Random Forest and SVM, Support Vector Machine was the most accurate predictor
- . The most important features in predicting crash severity are parked cars and number of cars.

## **CONCLUSION**

It will be hard to develop an application that predicts the severity of car crashes and alerts the drivers in real time as weather and other conditions that can be known in advance are not good predictors. As seen in the analysis, parked cars and numbers of vehicles are the most significant features, which requires the police to issue a detailed report right when the accident occurred and then send it to the application that will complete the prediction.

Therefore, to avoid car crashes, we cannot deploy a solution that can have a major impact, we can just help reduce impact / additional damage and thus drivers being careful remains the ultimate solution for avoiding car crashes.