# Capstone Project – Car Accident Severity: Seattle

NC

# Introduction

## Background

- Road traffic injuries cause considerable economic
- Road traffic accident caused most countries 3% of their GDP
- Hence it is important to analyse the severity of accidents and their causes

# Introduction

## Business Problem

- To predict the severity of accidents happening on the roads of Seattle based on multiple attributes

## Target Audience

- State Departments
- Drivers
- Logistic Companies
- Cab Companies etc.

# Data Acquisition and Cleaning

## Data Source

- Data is from Seattle Department of Transportation in CSV format
- The dataset is around 200,000 events

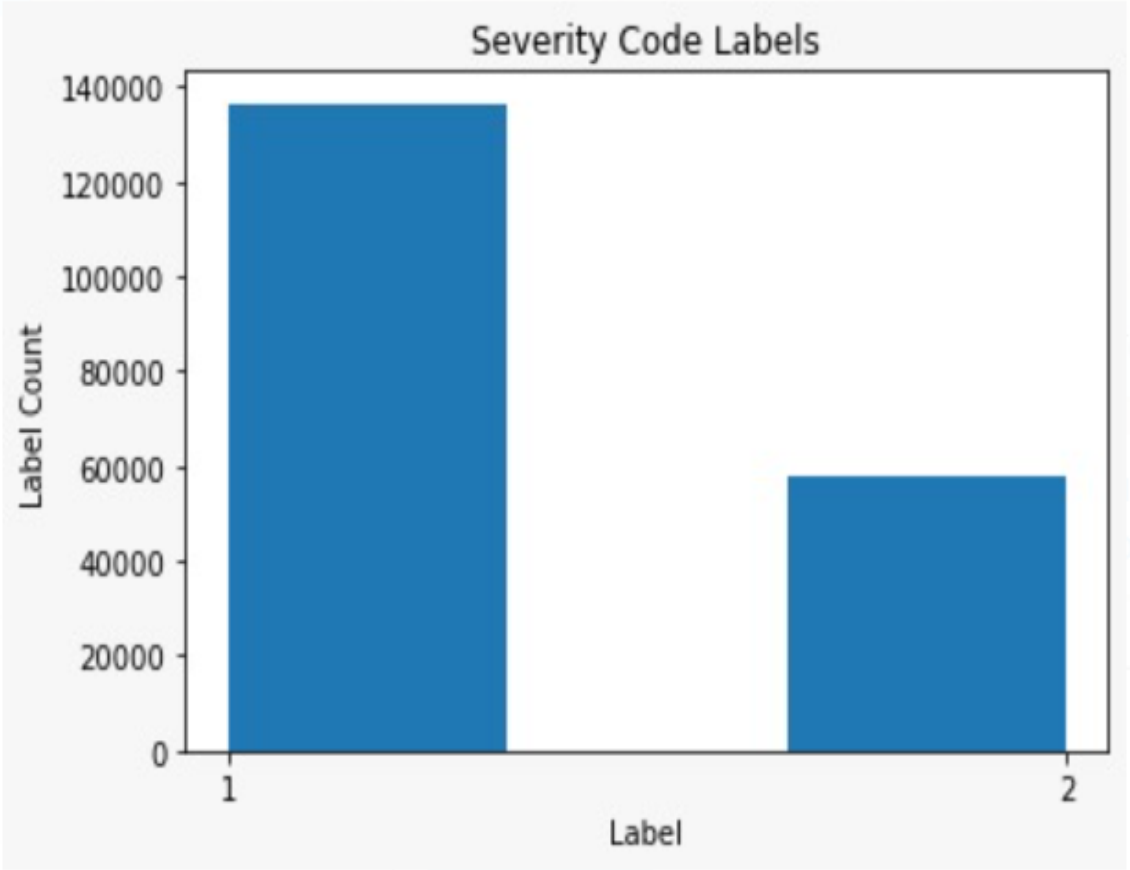
## Data Cleaning & Feature Selection

- Total 37 attributes
- 4 attributes were chosen based on negative correlation with the severity – WEATHER, ROADCOND, LIGHTCOND, VEHCOUNT

# Methodology

## Data Analysis & Preparation and Normalization

- Data Balancing
- Check & change the dtype
- Normalize the dataset
- Train Test Spilt



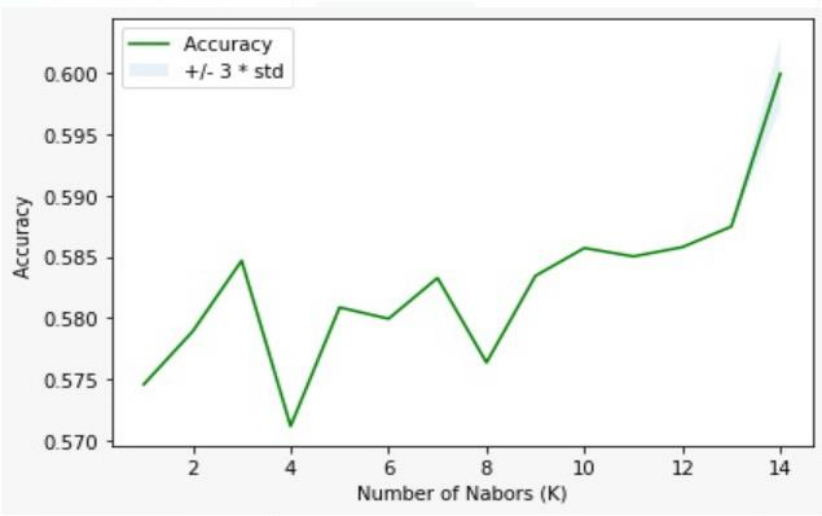
# Methodology

## Modelling/ Classification Source

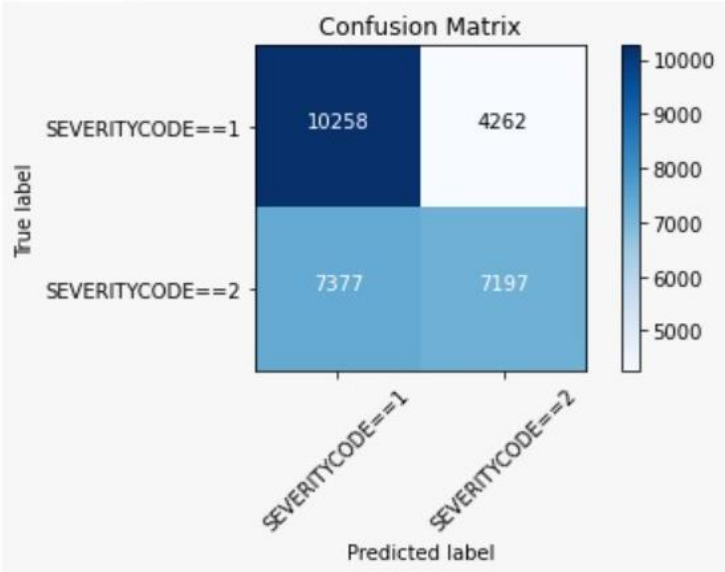
- K-Nearest Neighbours (KNN)

Train set Accuracy: 0.5995050525881626
Test set Accuracy: 0.599951880112738
The F1-score is 0.5953950705187826
The Jaccard similarity score is 0.4684659999086633

Best K = 14




Confusion Matrix for KNN



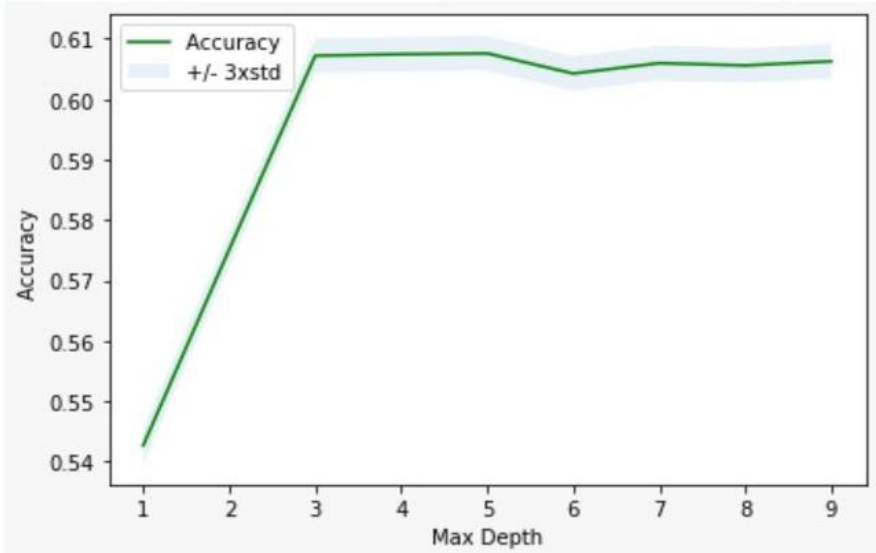
# Methodology

## Modelling/ Classification Source

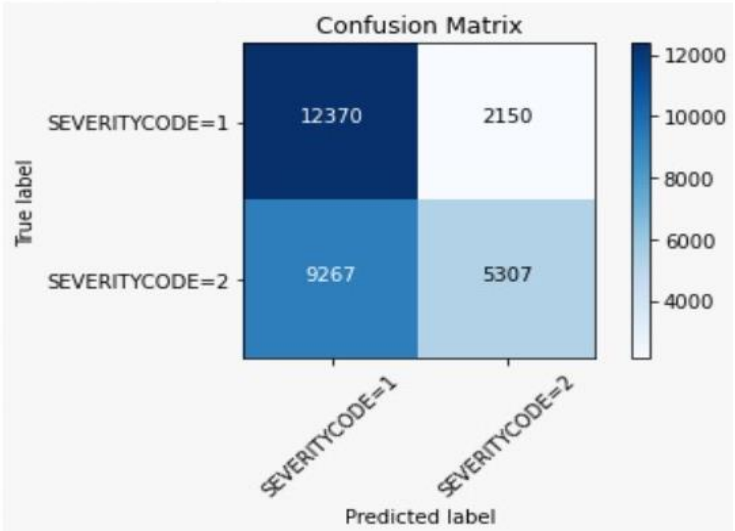
- Decision Tree (DT)

Train set Accuracy: 0.6059439517884558
Test set Accuracy: 0.607582319378566
DecisionTrees's Accuracy: 0.607582319378566
The F1-score is 0.5828190410528927
The Jaccard similarity score is 0.5200319502249128

Max Depth = 5




Confusion Matrix for DT



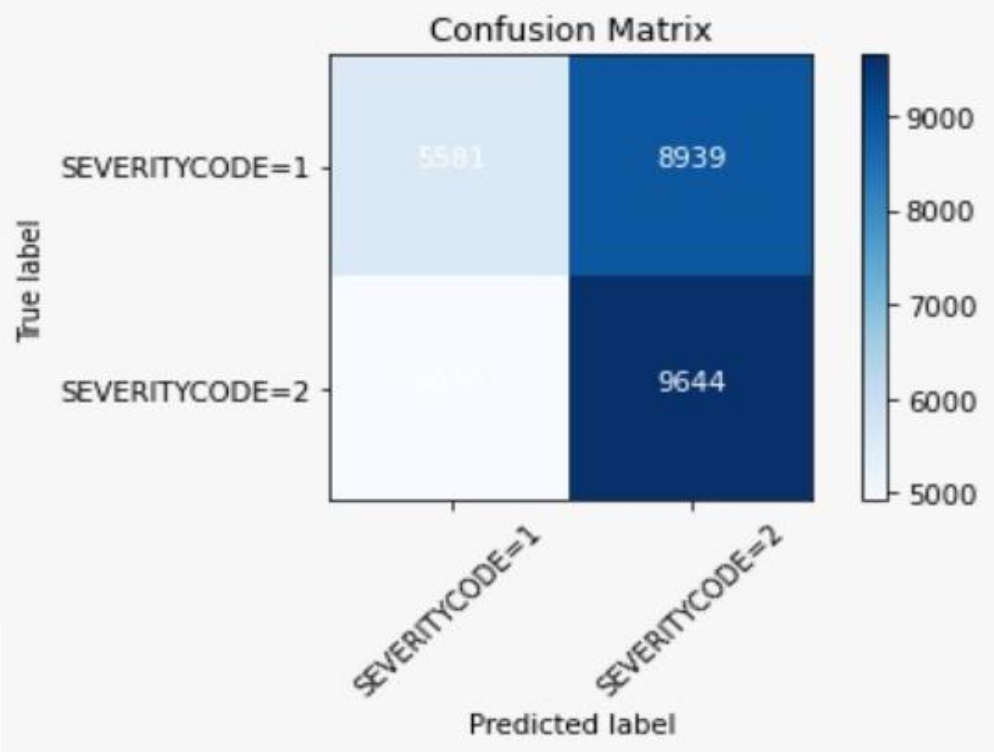
# Methodology

## Modelling/ Classification Source

- Logistic Regression (LR)

Train set Accuracy: 0.5257785110332027
Test set Accuracy: 0.5233037739740153
LR Accuracy: 0.5233037739740153
The F1-score is 0.5139481843676582
The Jaccard similarity score is 0.2869408740359897
The logLoss is: 0.6811191407676671

Confusion Matrix for LR



# Result and Discussion

- The best model: combination of high F1-score, high jaccard similarity score and the smallest log loss
- The best performance is observed in Decision Tree model

| | KNN | DT | LR |
|---|---|---|---|
| Accuracy | 0.5999 | 0.6075 | 0.5233 |
| F1-Score | 0.5953 | 0.5828 | 0.5139 |
| Jaccard similarity score | 0.4684 | 0.5200 | 0.2869 |
| Log Loss | - | - | 0.6811 |

# Conclusion

- To predict the severity of the car accidents that will happen in Seattle based on the known conditions like weather, lighting conditions, road condition and vehicle count
- Model can be used by individuals and many sectors
- Prepare them to handle the situation more effectively

# Thank you