# Prediction of Car Accident Severity

## Capstone Project

by

Iziegbe Harriet Igbinedion

# Introduction : Business Problem

- The aim of this project is to design a model to predict the probability of a severe car accident occurring on a road given the right data.

- The stakeholders interested in this project are the drivers and pedestrians who pass through the affected roads at a given time.

- We would use some data science tools and methodologies coupled with some machine learning algorithms to design, train, and evaluate our model.

# Data Understanding

The dataset has 194673 rows and 38 columns including Location, Severity Code, Weather, Road Condition etc. and this is made available in a CSV format.

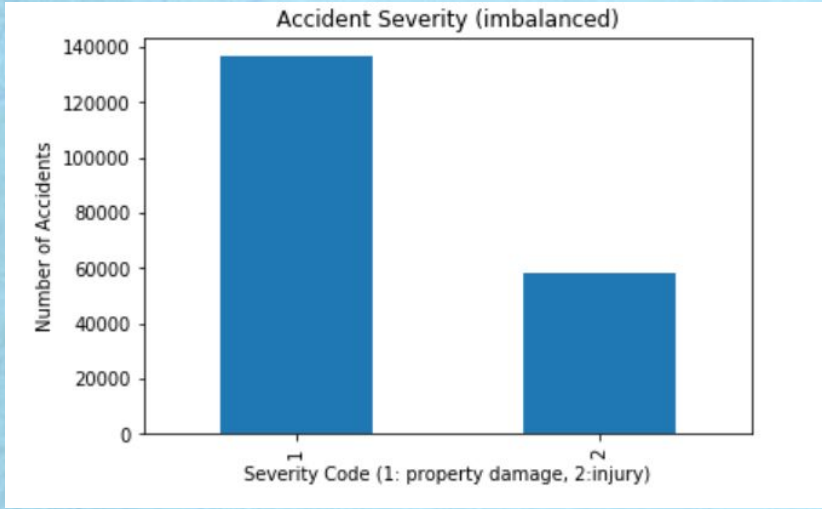
Our predictor or target variable is 'SEVERITYCODE' because this is a measure of the severity of an accident from 0 to 5 within the dataset.

We will use 'WEATHER', 'ROADCOND' and 'LIGHTCOND' to determine the severity of an accident.

# First 5 rows of the dataset (df.head())

| SEVERITYCODE | X | Y | OBJECTID | INCKEY | COLDETKEY | REPORTNO | STATUS | ADDRTYPE | INTKEY | ... | ROADCOND | LIGHTCON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | -122.323148 | 47.703140 | 1 | 1307 | 1307 | 3502005 | Matched | Intersection | 37475.0 | ... | Wet | Dayligh |
| 1 | -122.347294 | 47.647172 | 2 | 52200 | 52200 | 2607959 | Matched | Block | NaN | ... | Wet | Dark - Stree Lights O |
| 1 | -122.334540 | 47.607871 | 3 | 26700 | 26700 | 1482393 | Matched | Block | NaN | ... | Dry | Dayligh |
| 1 | -122.334803 | 47.604803 | 4 | 1144 | 1144 | 3503937 | Matched | Block | NaN | ... | Dry | Dayligh |
| 2 | -122.306426 | 47.545739 | 5 | 17700 | 17700 | 1807429 | Matched | Intersection | 34387.0 | ... | Wet | Dayligh |

# Oversampling & Undersampling




# Model Development

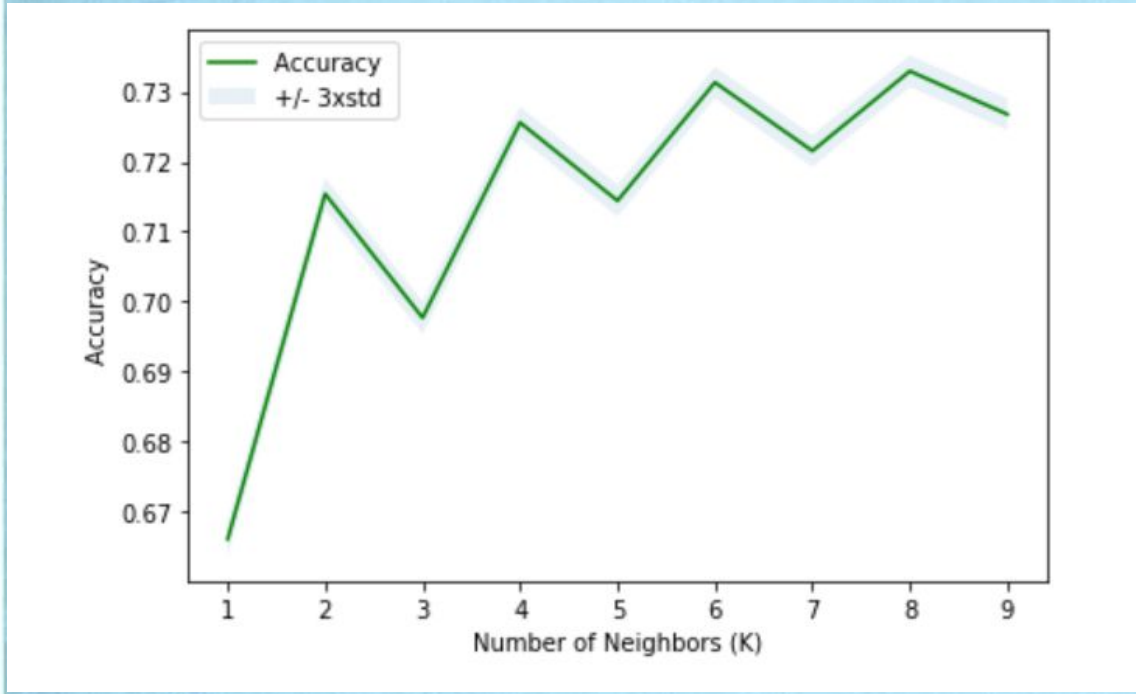
**MACHINE LEARNING ALGORITHMS:**

**K Nearest Neighbor**

**Decision Tree**

**Logical regression**

# K Nearest Neighbor



# KNN Evaluation

```
from sklearn.metrics import f1_score
f1_score(y_test, yhat, average='weighted')
```

```
0.7062731513289651
```

```
from sklearn.metrics import jaccard_similarity_score
jaccard_similarity_score(y_test, yhat)
```

```
C:\Users\izzie\anaconda3\lib\site-packages\sklearn\metri
has been deprecated and replaced with jaccard_score. It
behavior for binary and multiclass classification tasks.
  FutureWarning)
```

```
0.7030949017593425
```

# Decision Tree

```
print("DecisionTrees's Accuracy: ", metrics.accuracy_score(y_test, yhat))
```

```
DecisionTrees's Accuracy:  0.7429562090663927
```

```
print('f1 score:', f1_score(y_test, yhat, average='weighted'))
```

```
f1 score: 0.7002101557946572
```

```
print ('jaccard similarity:', jaccard_similarity_score(y_test, yhat))
```

```
jaccard similarity: 0.7429562090663927
```

# Logistic Regression

```
print("Logistic Regresion's Accuracy: ", metrics.accuracy_score(y_test, yhat))
```

```
Logistic Regresion's Accuracy:  0.7030949017593425
```

```
print('f1 score:', f1_score(y_test, yhat, average='weighted'))
print ('jaccard similarity:', jaccard_similarity_score(y_test, yhat))
```

```
f1 score: 0.6130193709022662
jaccard similarity: 0.7030949017593425
```

# Summary

From the results below, decision tree has proven to the best algorithm for predicting severity of car accidents with an accuracy score of 74%.

| | Algorithm | Jaccard | F1-Score | Accuracy | Logloss |
|---|---|---|---|---|---|
| 0 | KNN | 0.7267753948889174 | 0.7062731513289651 | 0.7255939386156415 | NA |
| 1 | DecisionTree | 0.7429562090663927 | 0.7002101557946572 | 0.7429562090663927 | NA |
| 2 | LogisticRegression | 0.7030949017593425 | 0.6130193709022662 | 0.7030949017593425 | 24.328955222549084 |

# Conclusion

**From the data set provided to us, we can see that change in weather is being directly linked to some classes.**

**From this statement, we arrived at the conclusion that a change in weather can, to an extent can affect one's mode of travel and like-hood of code 1 or code 2 damages occurring (property loss and injuries respectively).**