# Seattle Car Accidents Analysis

## Introduction/Business Problem

As the world is growing tremendously,demands of the people also increased exponentially. Technology is one of the most important driving factor behind the growth. For ease of transportation, hybrid /electric cars were been introduced to the market which attracts the majority of the population.Therefore,more vehicles on the road leds to the drastic leap in the collision rates as well.

The Seattle government is going to prevent avoidable car accidents by employing methods that alert drivers through alert display system and police to be more careful in critical situations.In most cases, not paying enough attention during driving, drugs and alcohol or driving at very high speed for fun/competition are the main causes of occurring accidents that can be prevented by deploying harsher regulations. Besides the causes, weather, visibility, or road conditions are the major uncontrollable factors that can be prevented by revealing hidden patterns in the data and announcing warning to the local government, police and drivers on the targeted roads.

The target audience of the project is local Seattle government, police, rescue groups and insurance organisation as well. This model results will provide them to make insightful decisions for reducing the number of accidents for the city.

## Data Set

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present. The data consists of 37 independent variables and 194,673 rows. The dependent variable, “SEVERITYCODE”, contains numbers that correspond to different levels of severity caused by an accident from 0 to 4. Severity codes are as follows:

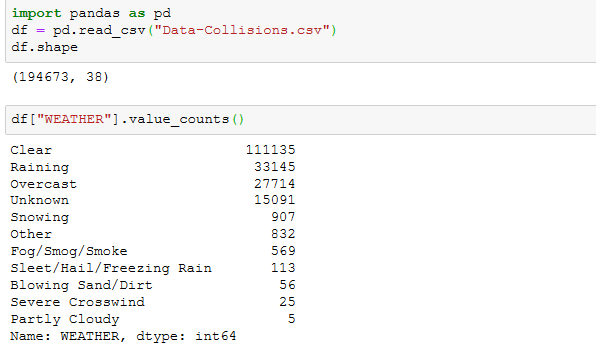
* 0: No Probability Clear Conditions)
* 1: Very Low Probability — Chance or Property Damage
* 2: Low Probability — Chance of Injury
* 3: Mild Probability — Chance of Serious Injury
* 4: High Probability — Chance of Fatality

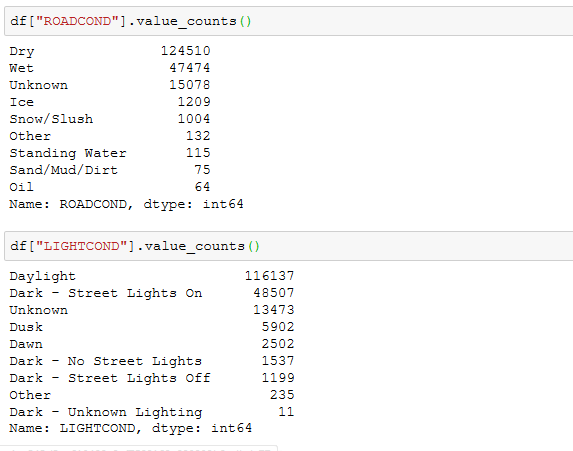
Furthermore, because of the existence of null values in some records, the data needs to be pre-processed before any further processing. Among all the features, I believe the following features have the most influence in the accuracy of the predictions:

* “WEATHER”,
* “ROADCOND”,
* “LIGHTCOND”

The target variable is “SEVERITYCODE”.

Please find the category wise results in case of weather/Road Cond/Light Cond as below.

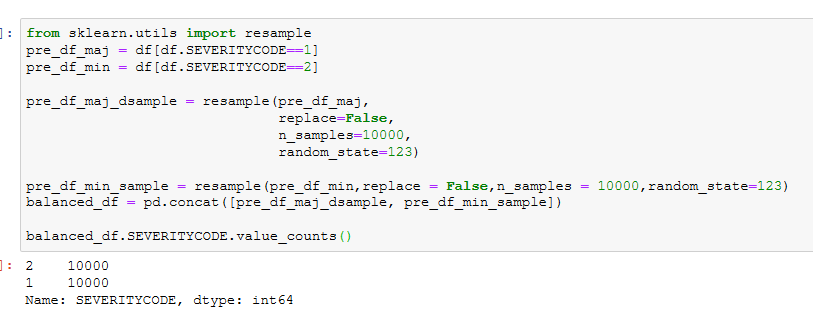




## Data Preprocessing

Data is in raw format and in order to evaluate the data we need to convert categorical variables as Boolean variables. After analysing the data, I decided to focus on three main independent variable “Weather”, “Road Conditions”, “Light Conditions” to predict the target variable severity.

Downsampling has been done to balance the data in the given dataset.



Feature extraction has been done for the independent variables:

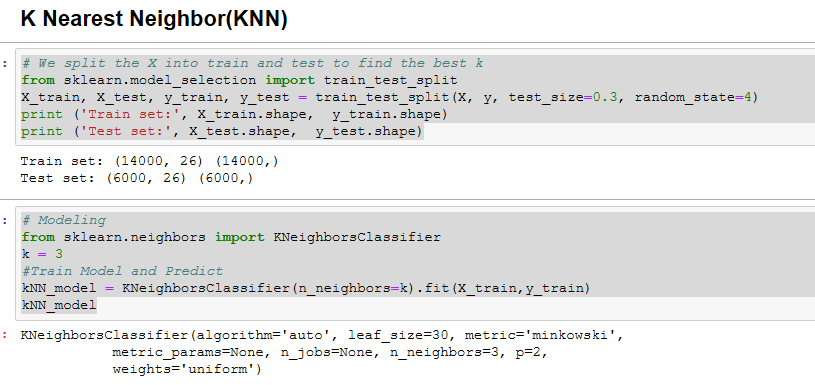


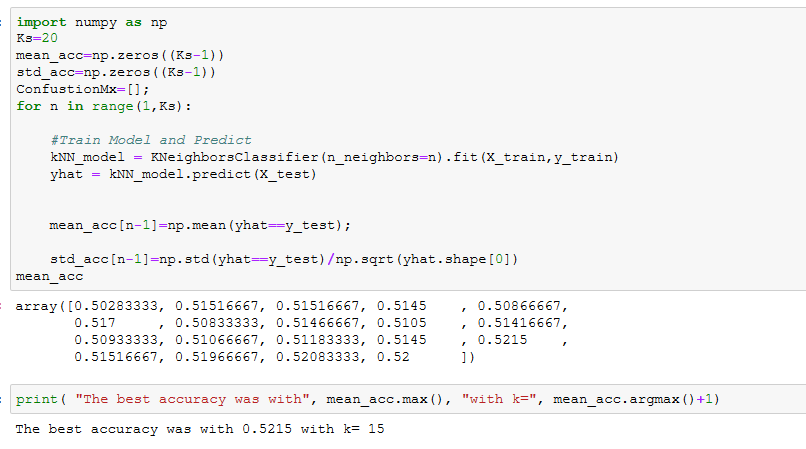
## Methodology

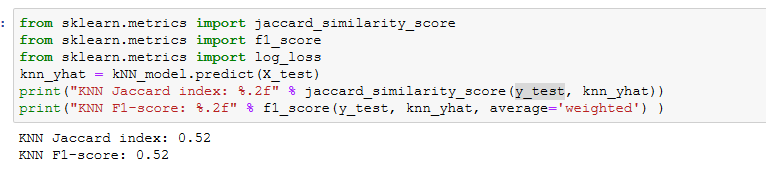
After balancing the data and feature extraction, I had applied the machine learning models under classification techniques. Models are as follows:

* 1. K Nearest Neighbour
  2. Decision Tree
  3. Linear Regression

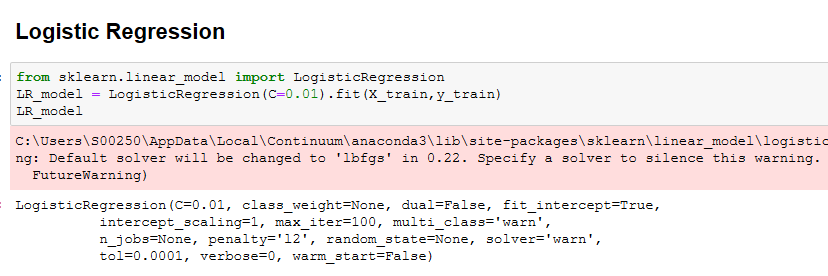
After importing required packages and splitting the data set into train test model. I had evaluated every model and results shown as follow:

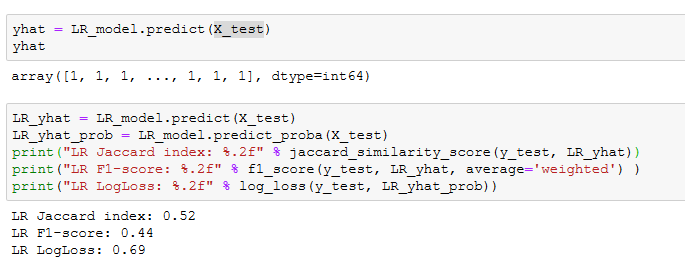












## Results and Evaluation

The final results of the model evaluations are summarized as in the following table:

|  |  |  |
| --- | --- | --- |
| ML Model | Jaccard Score | F1 Score |
| KNN | 0.52 | 0.52 |
| Decision Tree | 0.52 | 0.46 |
| Logistic Regression | 0.52 | 0.44 |

Based on the above table, KNN is the best model to predict car accidents severity.

## Conclusion

Based on the Seattle Car accidents analysis, we can infer most of the severity class 1 /2 cases occurs due to the Weather, Road and Light Conditions.