Car Accident
IBM Capston Final Project

MARIO PAREDES URIONA MSC
SEPTEMBER, 2020

1.1 Background

Car accident is one of the most common accident in the world, especially in crowded cities. There are different factors that are involved in this kind of accident, such as weather conditions, road conditions light conditions and more. The mix of these factors can obtain a higher probability to get an accident. Thus, it is important to analysis all the conditions to calculate this probability.

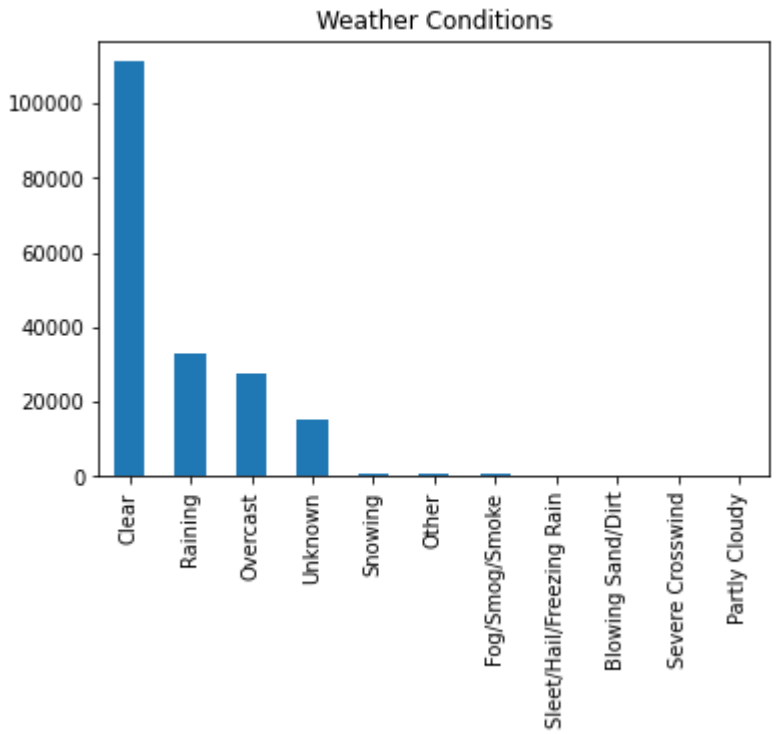
1.2 Problem

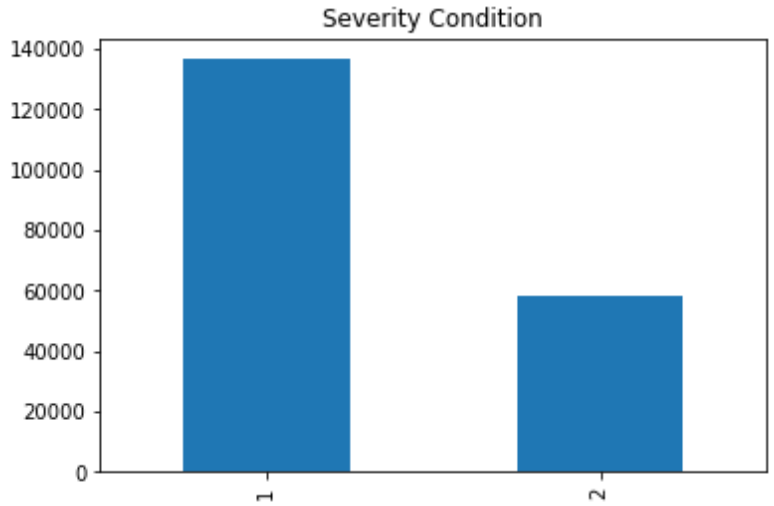
The number of car accident in Seattle is noticeable. Thus it is important to determine the probability to get an accident depending of the different conditions of the road, day and more.

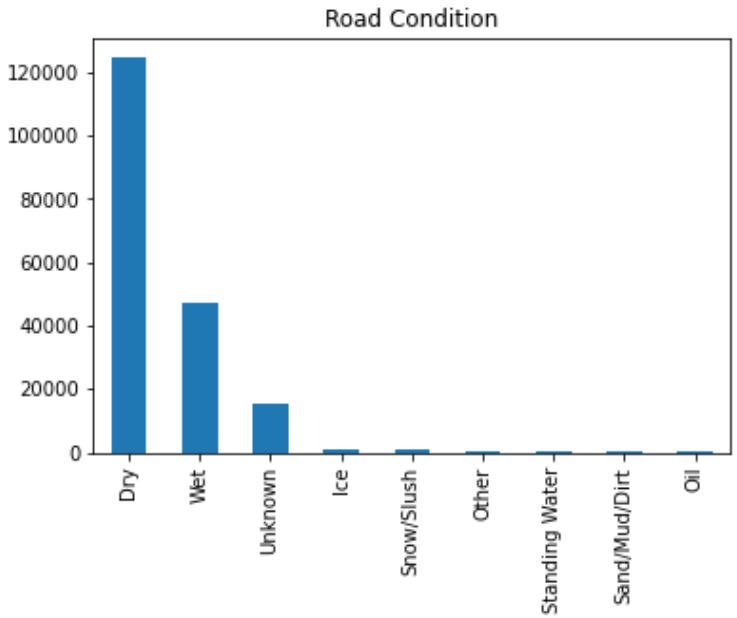
Data Sources

Data acquired from the historical data from Seattle. this data base includes different attributes, such as location, date of accident, severity of the accident and more. 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.

Exploratory Data Analysis

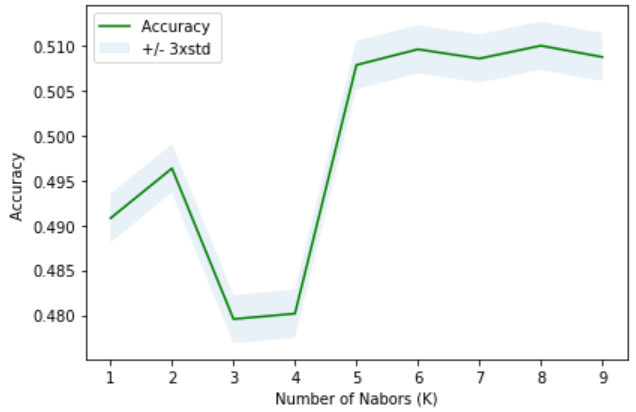






K-Nearest Neighbors

The k-nearest neighbors (KNN) algorithm from supervised machine learning algorithm that can be used to solve both classification and regression problems.



Results

Train Accuracy	0.50916
Test Accuracy	0.50960
F-1 Score	0.49953
Jaccard-Score	0.27193

Decision Tree

Results

Decision Tree Accuracy	0.52533
F-1 Score	0.49260
Jaccard-Score	0.26872

Logistic regression

Decision Tree Accuracy	0.52198
F-1 Score	0.51599
Jaccard-Score	0.40015

Final Results

	KNN	Decision Tree	Logistic regression
Model Accuracy	0.50960	0.52533	0.52198
F-1 Score	0.49953	0.49260	0.51599
Jaccard-Score	0.27193	0.26872	0.40015

Conclusions

In this works, different classification methods were applied into a set of car accident data from Seattle in order to obtain the probability to suffer a car accident depending of different factor such as weather condition, road condition and light conditions. these parameters were defined as the main parameter for this study.

As a result, the algorithm to measure the probability to suffer an accident was developed with a considerable good accuracy.