Analyzing and Predicting accident Severity.

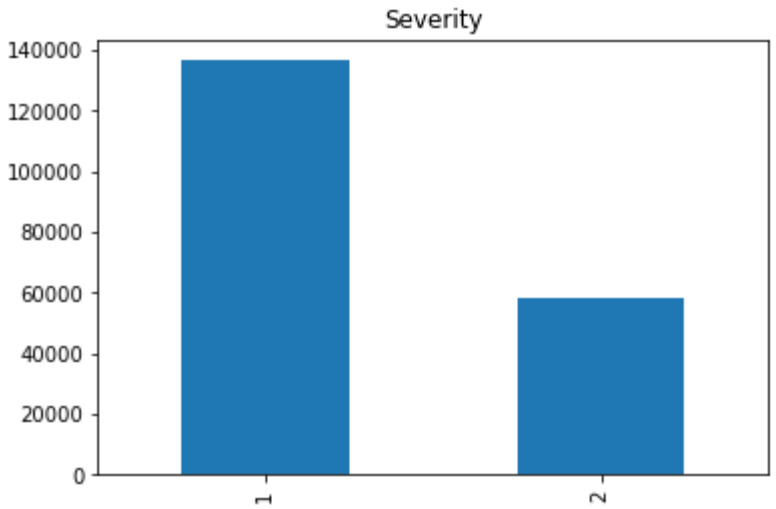
Background

- Traffic Data is available for the period from 2014
- It is required to analyze the factors leading to accidents.
- There are two type of accidents Severity based on fatality and property damage.
- There are several information in the data set, We considered Weather, Road Condition, Light Condition.

Data

- Traffic Data is available for the period from 2014
- There are 194,672 data with 37 different attributes.
- It is required to analyze the factors leading to accidents.
- There are two type of accidents Severity based on fatality and property damage.
- There are several information in the data set, We considered Weather, Road Condition, Light Condition.
- Data is cleaned and normalized for prediction.
- Most of the interested data are categorical variables , that are converted to numeric for model building.

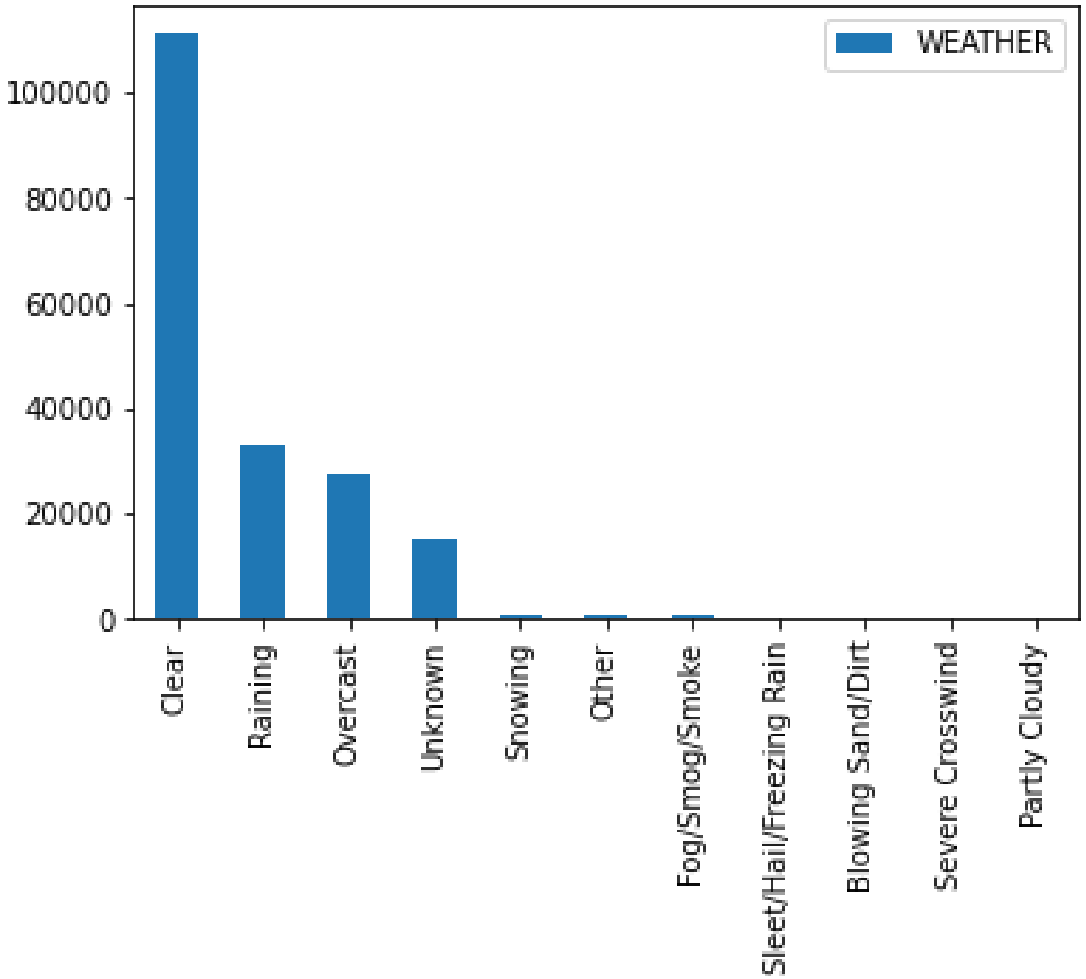
Severity of Accidents



	SEVERITYCODE
1	136485
2	58188

Most of the accidents are Severity 1- Property Damage only

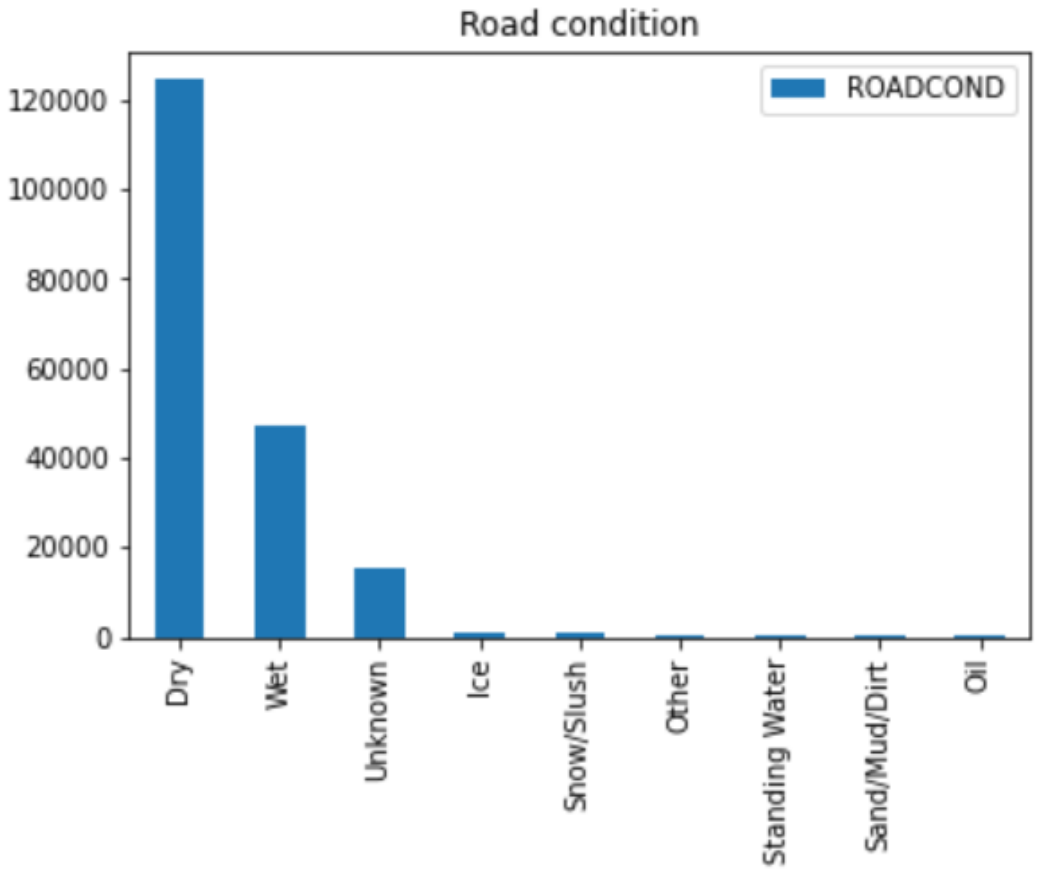
Weather Condition and Accidents



	WEATHER
Clear	111135
Raining	33145
Overcast	27714
Unknown	15091
Snowing	907
Other	832
Fog/Smog/Smoke	569
Sleet/Hail/Freezing Rain	113
Blowing Sand/Dirt	56
Severe Crosswind	25
Partly Cloudy	5

- Most of the accidents are happening clear weather condition
- Meanwhile, Raining is the second highest and it could be one of the reason for accidents.

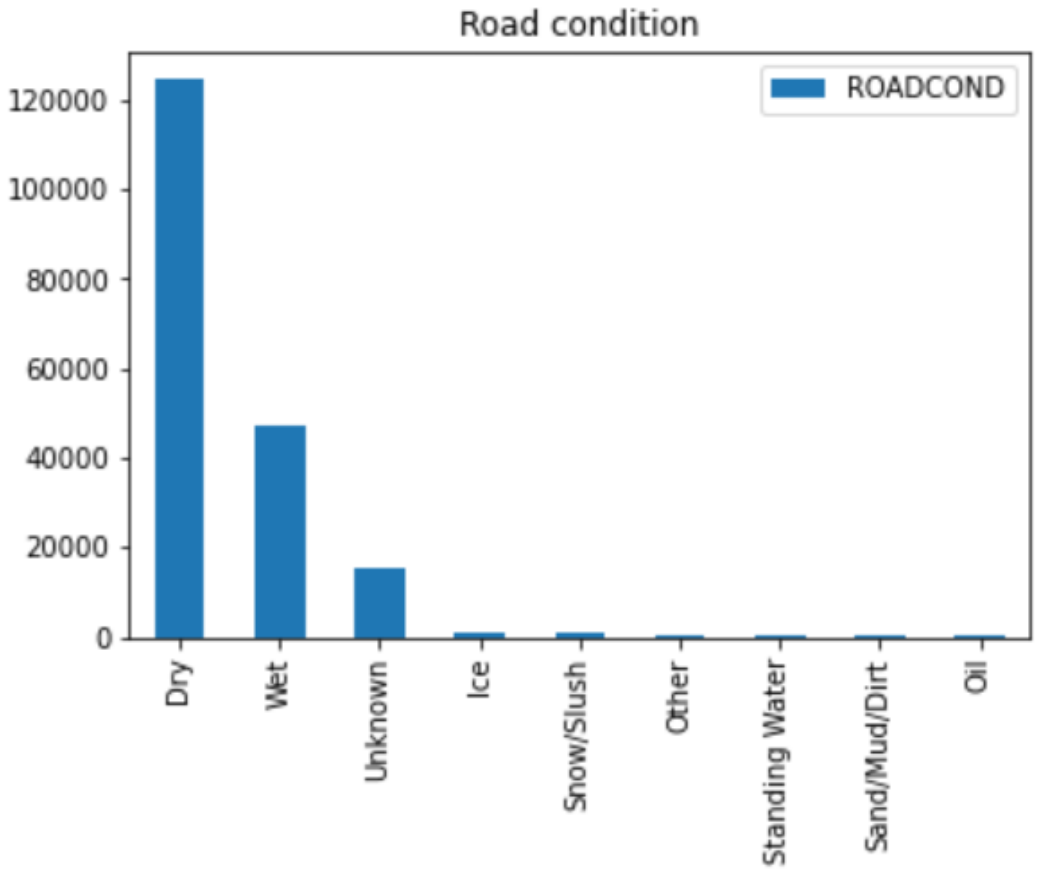
Road Condition and Accidents



	ROADCOND
Dry	124510
Wet	47474
Unknown	15078
Ice	1209
Snow/Slush	1004
Other	132
Standing Water	115
Sand/Mud/Dirt	75
Oil	64

Most of the accidents are happening normal road condition – Dry.

Road Condition and Accidents

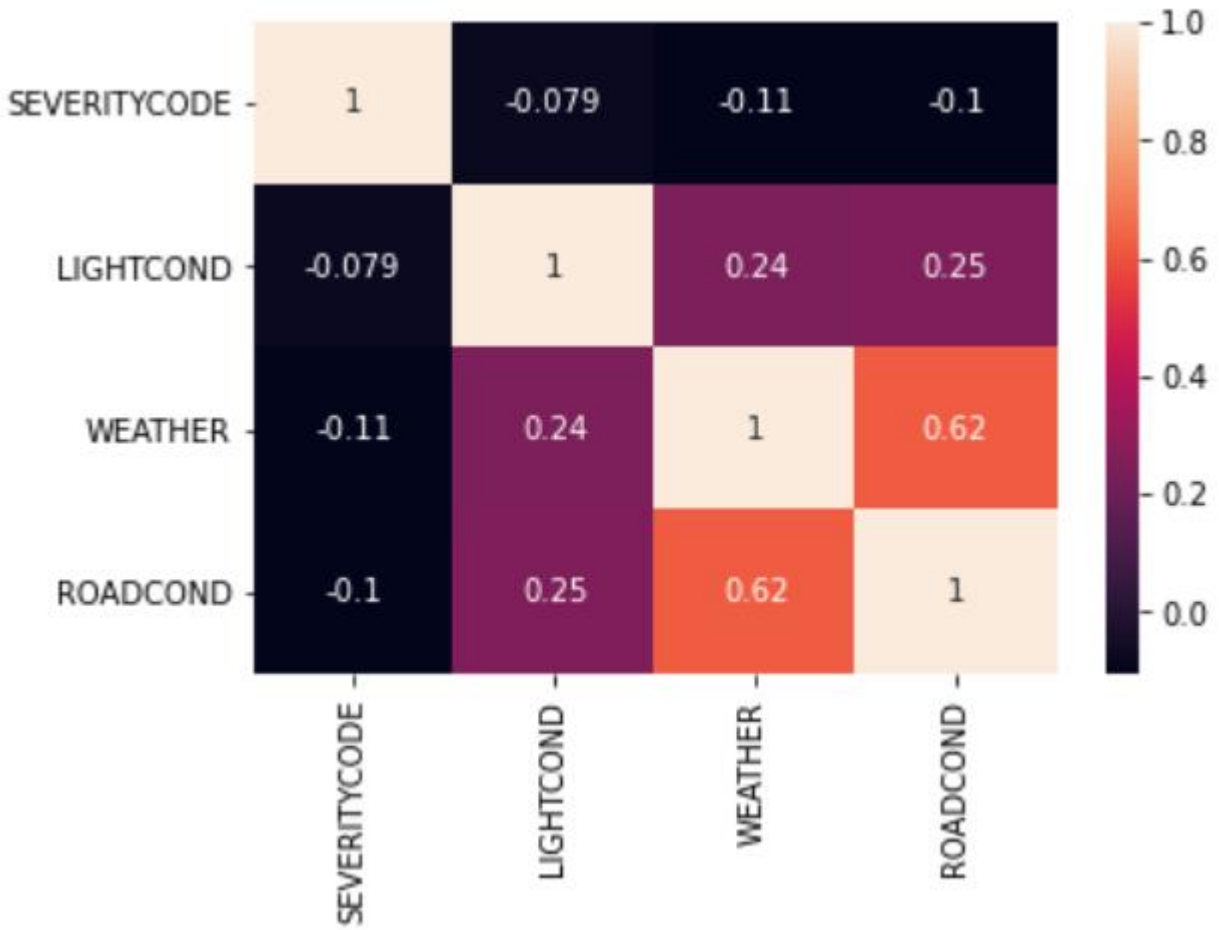


	ROADCOND
Dry	124510
Wet	47474
Unknown	15078
Ice	1209
Snow/Slush	1004
Other	132
Standing Water	115
Sand/Mud/Dirt	75
Oil	64

Most of the accidents are happening normal road condition – Dry.

Model Building

Given data split of



- Looking in the heatmap and correlation , Most of the values are less than .5 (50%).
- There is no significant correlation between Severity , Weather , Road Condition and Light condition.

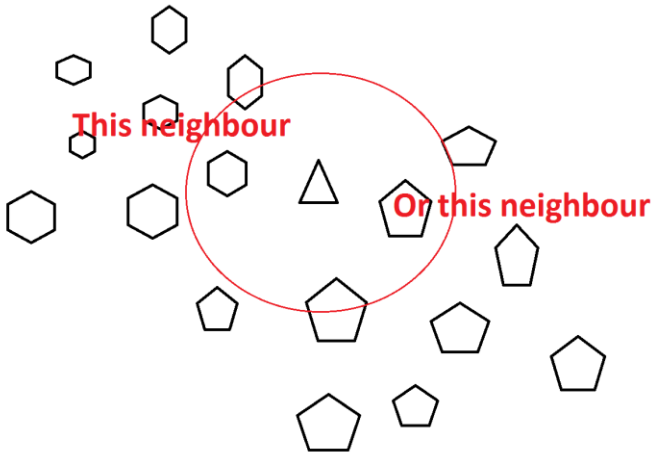
Model

- Model is build to predict the severity of accident (Severity 1 or 2)
- Data is normalized (Down sampled) to match the count as most of the data is severity 1 data.
- Predicted model is used for 70% of normalized data and 30% data is being evaluated
- 50% to 60% of data accuracy found in the prediction.
- 3 algorithms are used to build different models.
- Model build will be evaluated and will be adjusted in future based on the output.

K-Nearest Neighbors (KNN)

Build the learning model based on K-Nearest Neighbors (KNN)

Severity of an accident will be decided based on historical data point on how closer/neighbor is classified for independent variables, that is Weather, Road condition and Light condition.

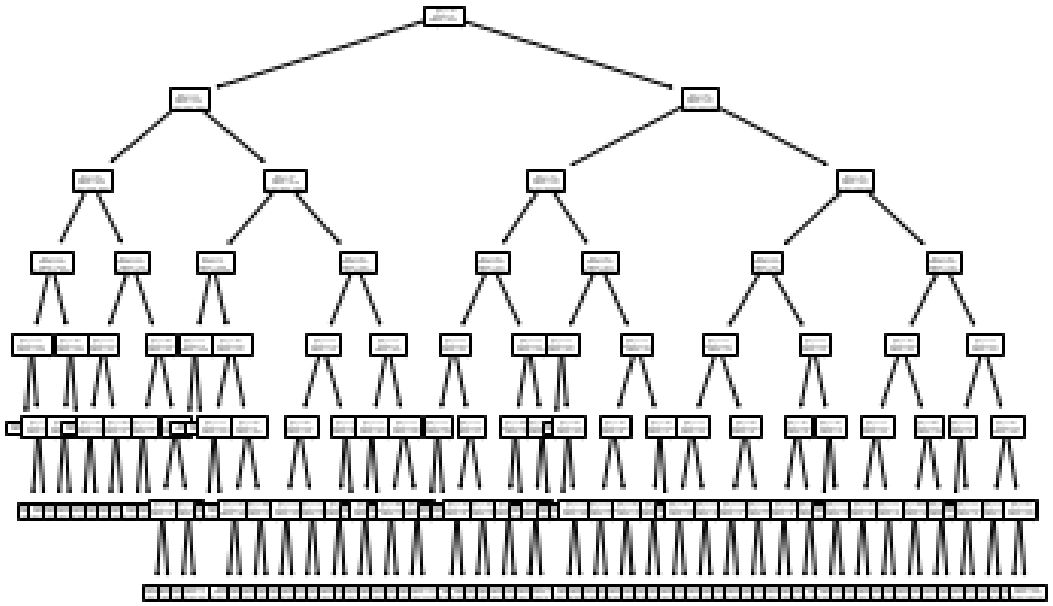


- Severity is predicted based on 21 nearest neighbors.
- Found it is 55% accurate.
- It is noted the accuracy level is very low.

Decision Tree

Build the learning model based on Decision Tree.

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences. We are predicting the severity by building the tree based on Weather, Road condition and Light condition.



- Found it is 56% accurate.
- It is noted the accuracy level is very low.

Logistic Regression

Build the learning model based on Logistic Regression

Logistic regression is one the best model to be used for as pass/fail, win/lose, alive/dead or healthy/sick; In our case Severity 1 or 2

- Found it is 50% accurate.
- It is noted the accuracy level is very low.

Conclusion and Future Direction

- Model build to be evaluated for some period of time.
- Traffic information can be made available for future analysis.
- Weather and Light condition is having impact on the number of accdidents.
- Accidents severity for Dark light and Rainy condition to be further Analyzed.

Thank You