

Predicting Car Accident Severity

2020

09/21/2020

Authored by: Shivani Tare

Tabel of content	
1.0 Abstract -	3
2.0 Introduction	3
3.0 Data Used and visualization	4
4.0 Methodology	8
5.0 Conclusion	9

1.0 Abstract

The problem present here with us is with increase in use of cars and newer technologies that are coming up the number of accidents is getting multifold. According to certain reports Seattel city of US is amongst the tops cities in number of accidents in world. In these accidents many precious lives are lost and property worth millions are damaged. We need to find a way if we can reduce number of accidents and with that reduce the loss of lives and property. We have the data for the accidents that have occurred in the city from year 2004 with various attributes assigned to it such as when the accident occurred date-time what were the road, light and weather conditions, was the drives under the influence of alcohol, did the accident result in property damage only or there were Injuries to people.

From this data we can think of predicting the severity of accident if it will occur would it result in property damage or Injuries and how sever it would be. We can use this to restrict speed limit on certain days when there are high chances of sever accident. We can use this to transmit messages across the city that there are probabilities that an accident can occur please drive safely. And sign boards where there are more accident happening – an Accident-prone zone. Basically, we can apply stricter laws and take various precautions so that the numbers of accident are reduced.

We will create a machine learning model to perform our task of predicting accident severity based on various conditions.

2.0 Introduction

a. Business Problem

As stated earlier the business problem here in front of us is to predict severity of accident before it happens based on various conditions that are there. The conditions which we can use to predict severity could be road, light, weather conditions, if the driver is under the influence of alcohol or not if the day of week makes any impact on severity of accident etc.

b. Problem

Data for the accident might be coming from various source like police records, hospitals were the patients involved in accident were being treated. Transport department etc. The project here aims to predict the severity of the accident based on the data that we already have. We will need to clean and transform the data before we use it to train our model and predict the severity of accident.

c. Interested

State and transport department would be interested so that based on the condition they could decide the speed limit, transmit the message that condition are such that you could meet an accident please drive safely or even keep a marking at points where frequently sever accidents take place.

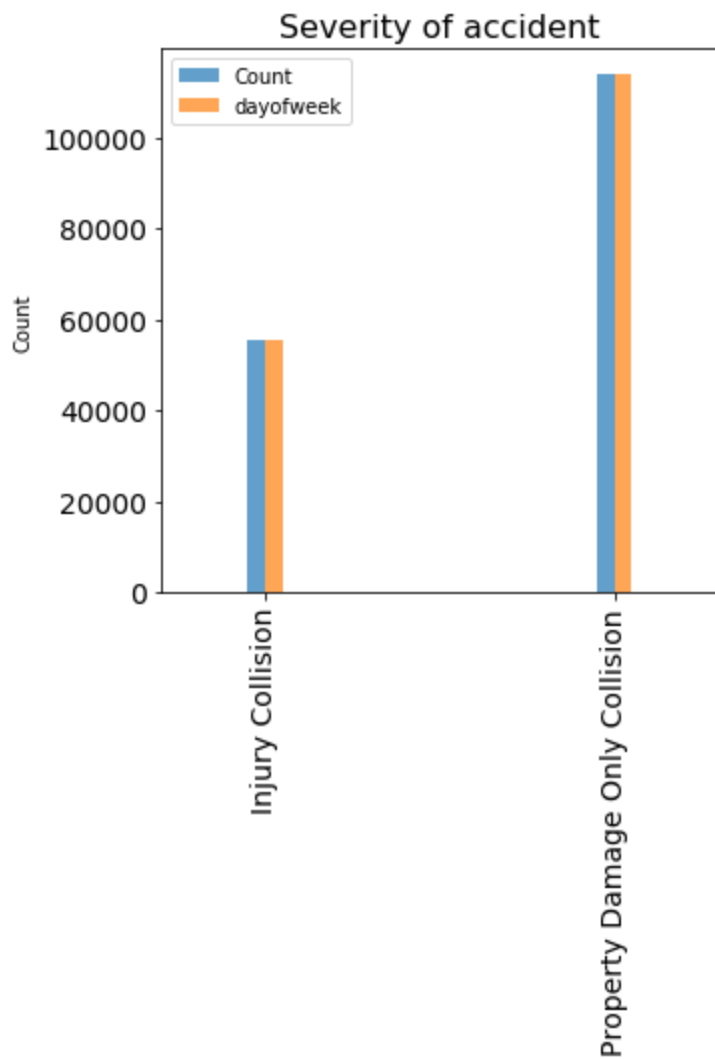
3.0 Data Used and visualization

Data that I have used here is that which is shared with us. It contains accident data from Seatttel city. There are various columns which are redundant and meaning less with respect to the data model which we are going to create so we would drop those columns. Some of columns that are drop are OBJECTID, INCKEY, COLDETKEY, REPORTNO and JUNCTIONTYPE.

On visualizing various columns, we can understand the data better below are various graphs from which we can understand the data.

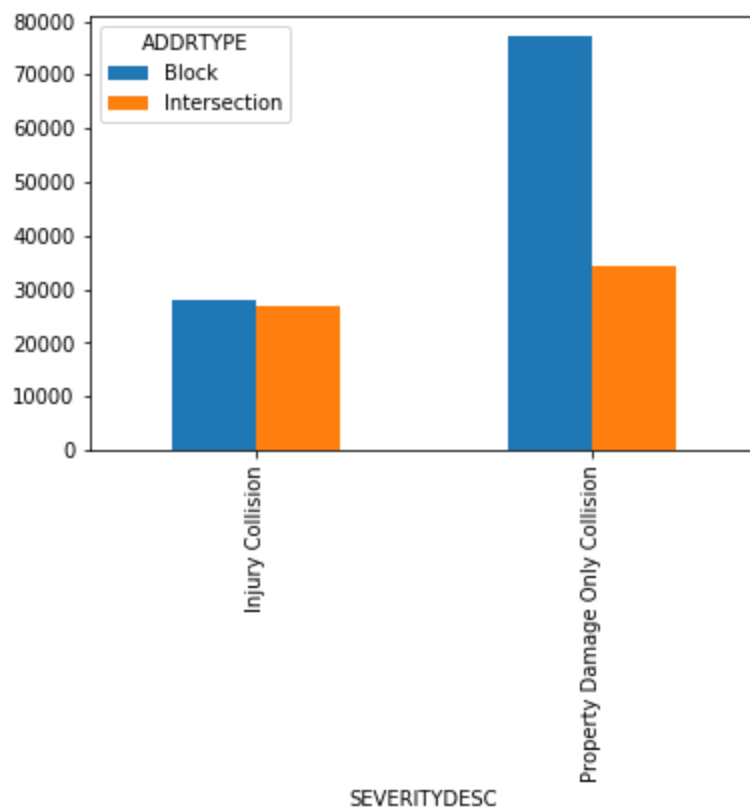
1. Severity of accident

	SEVERITYDESC	Count
0	Injury Collision	55539
1	Property Damage Only Collision	113700



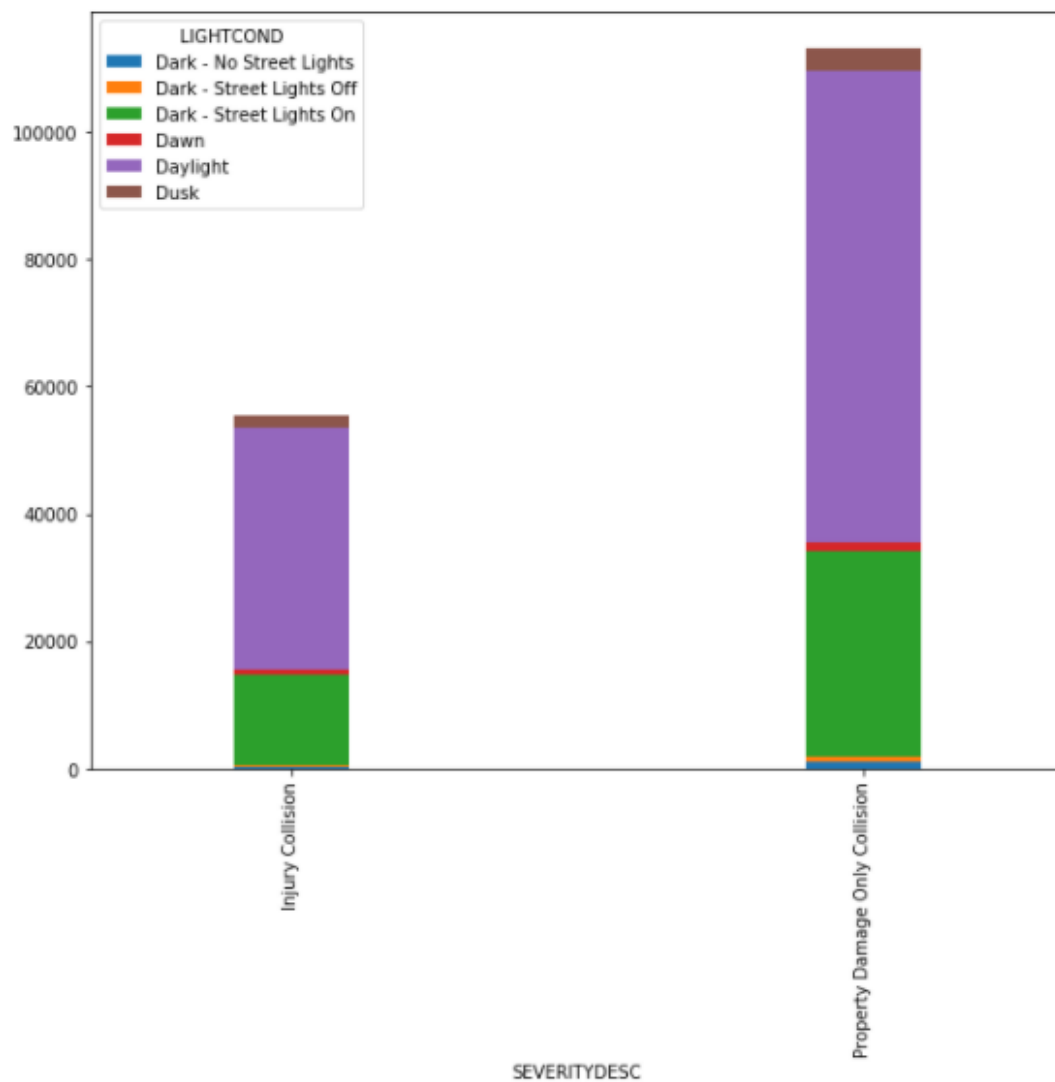
From the above graph we can clearly understand that the Property Damage Only Collision is more and Injury collision so while we are creating a model, we will have to downsize the number of rows for property damage only collision otherwise it will create a bias in the model.

2. Based on junction type if it is Block or intersection of two roads.



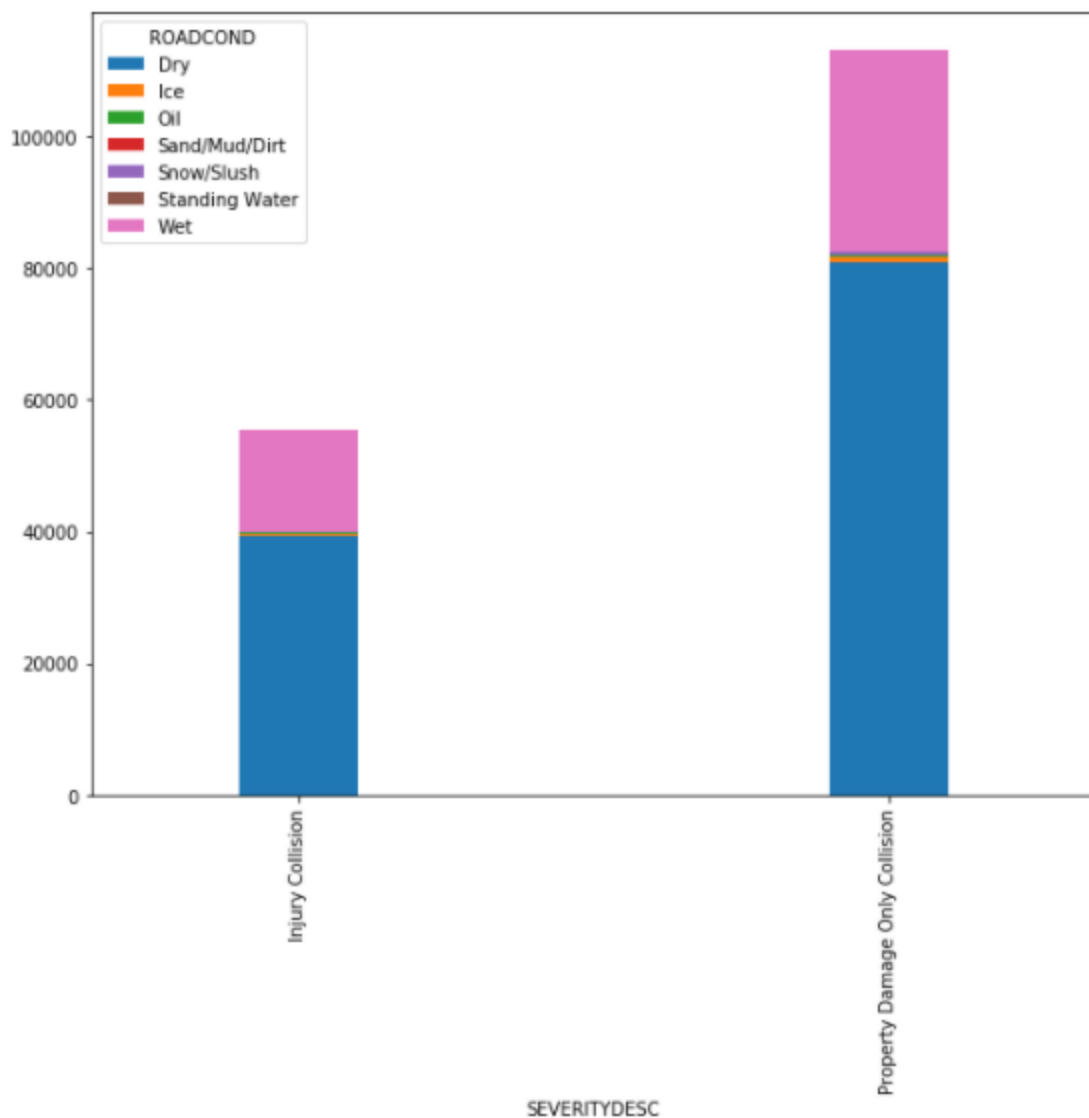
From the above graph we can clearly understand that more accident occurred at block.

3. Now let's see how light condition affects accident severity.



From the graph we can clearly understand that majority of accident occurred in Daylight or when it was dark with street light were on – this means the street light is the city are not proper and the states need to look into it.

3. Road Conditions



From the above data we can understand that majority accident occurred when the road where dry or either wet.

Similarly, we will visualize various data in the data set to understand the data.
(you can see the various data visualization in the Data visualization.ipynb notebook)

4.0 Methodology used

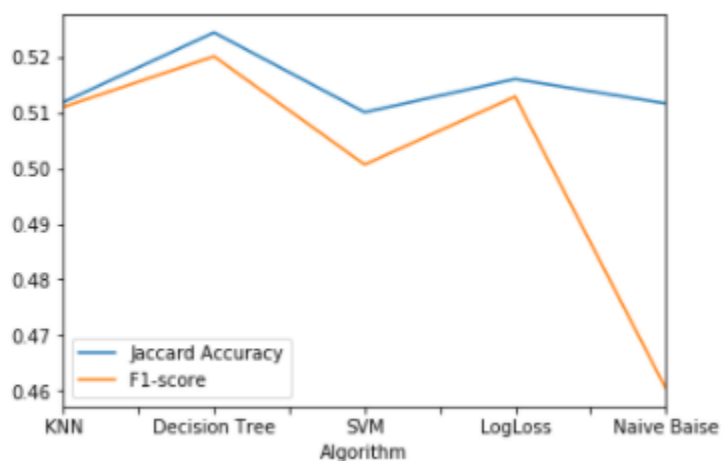
As we have labeled columns we would use supervised data modeling technique. And the data which we need to predict is SeverityCode – which is classified into two classes: one is Injury collision and another is property damaged collision, so we will use classification algorithm. We will use various supervised classification algorithms on the data provided and check which is the best algorithm for our car accident severity prediction.

The algorithms that I have used to see the accuracy are

1. K-Nearest Neighbor
2. Decision tree
3. SVM
4. LogLoss
5. Naïve Baise

Below is the graph of accuracy on F1-score and Jaccard similarity score

	Jaccard Accuracy	F1-score
Algorithm		
KNN	0.511960	0.511085
Decision Tree	0.524491	0.520196
SVM	0.510127	0.500702
LogLoss	0.516107	0.512959
Naive Baise	0.511719	0.460408



From the graph we can clearly see that the best fit algorithm for our data is Decision tree. And we will use it for prediction from now.

5.0 Conclusion

From above all Analysis, Visualisation and modeling we can conclude that it will be of great use if we can predict the accident severity by creating the model and deploying it so that everyone can use it, and benefit from it