Accident Prediction Report

Bikas Kar

Introduction

In this capstone we are going to solve the problem of finding the severity of the accident based on the weather condition or lighting some other conditions which can lead to accident. In this capstone we will also try to predict the probability of the accident which can be avoided people travelling on the road.

Background of this project is to find the causes of the accident and if we can figure out way to predict an accident and thereby saving precious lives. This prediction will help not only the commuters to be alert while driving, it can be used by traffic management officials to flash important messages using billboards.

Data Description

In this capstone we are going to use the data which is available in this site which contains the various attributes of each accident. there are 37 attributes for each accident. In this the most important one is the security code which is what we have to predict Based on other attributes . Some of the attributes which we Can use our location of the accident, lighting conditions, road conditions, Weather conditions, speed of the vehicle etc

We can create a model Which will predict of accident and also the probability and train the model using training data and then further test it with test data which will be a portion of the current data which we have.

After having a quick look at the data we can see that most of the accidents are either of severity code 2 or 1. For severity code 2, which is an injury most of the accident are happening at the intersection so this is also be a key factor 2 predict an accident.

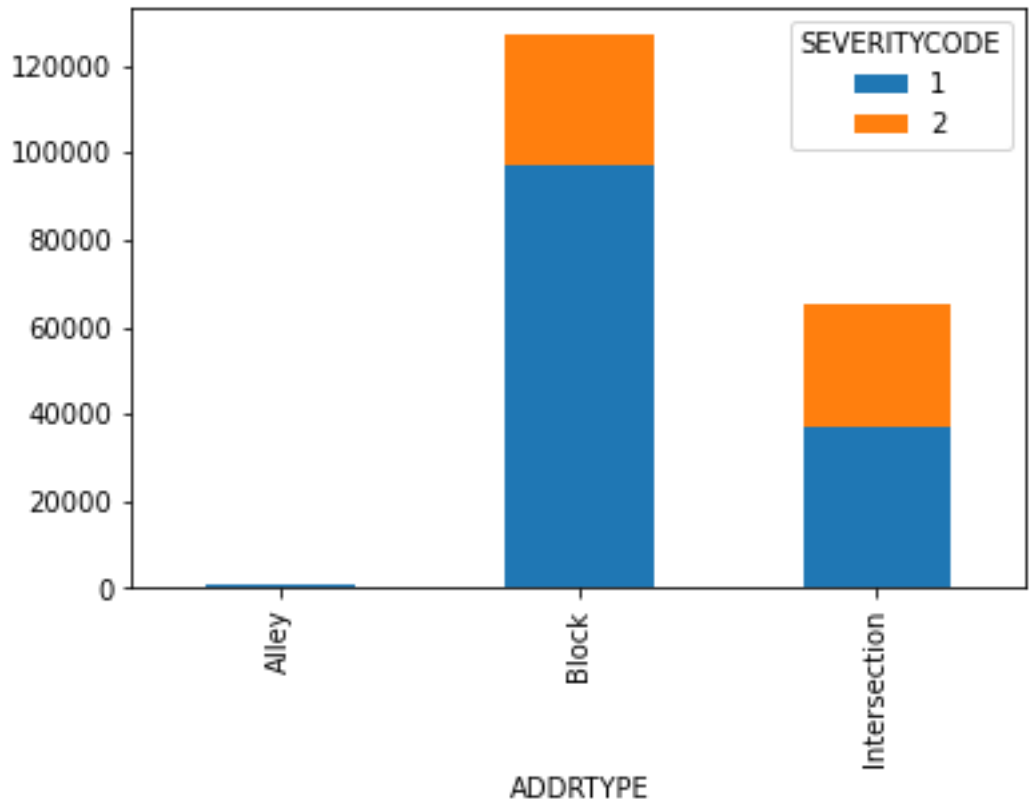
Also we can see that most of the accident at least 50% are happening when they weather condition is not clear. So as you we can have a insight in the current data and predict the severity of the accident

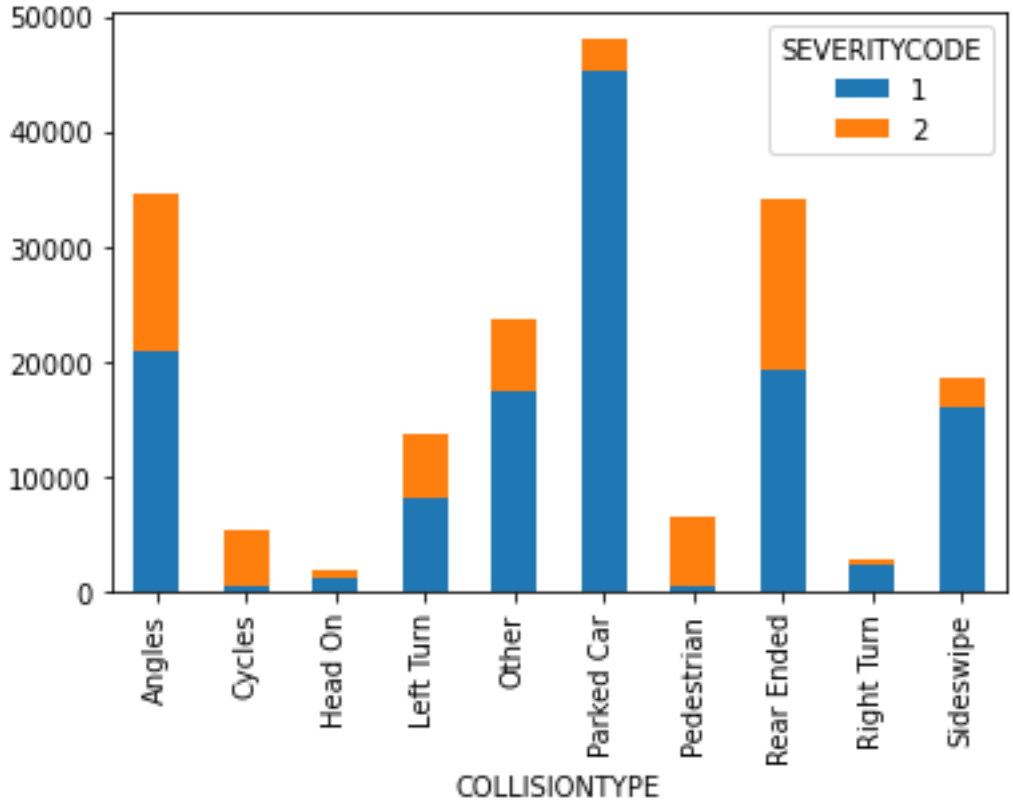
Data Analysis

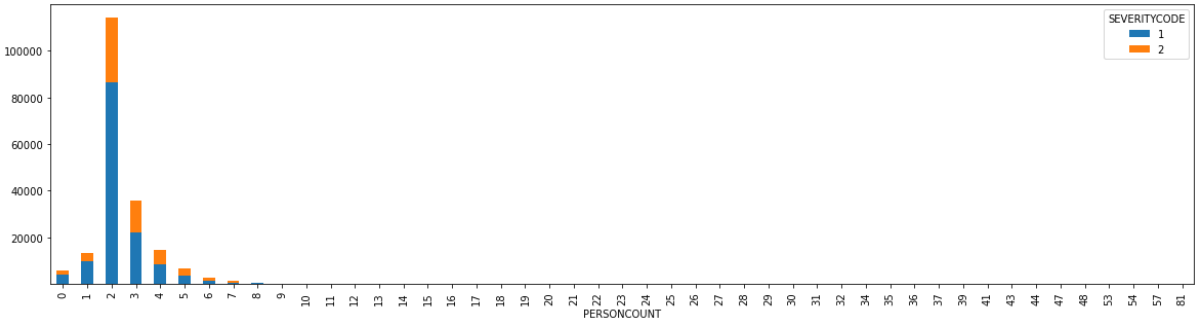
In this section, we will try to understand dataset and look into each column and its values. After this process, we will come to know which columns will be useful for us to use in our model.

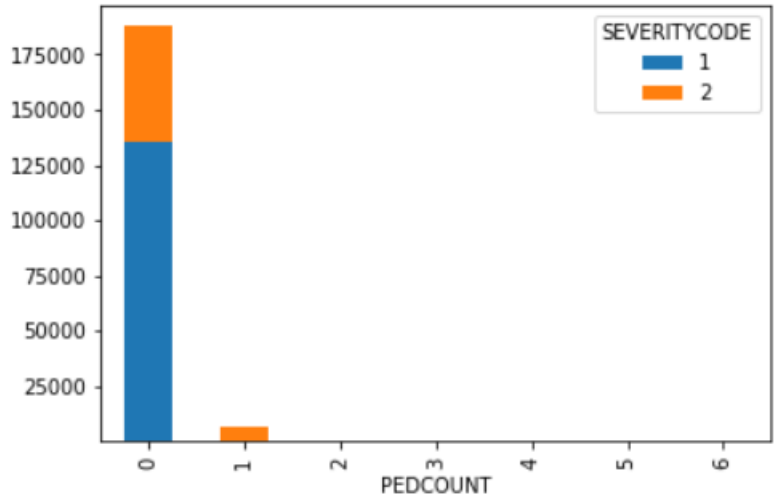
After looking at all the data, I arrived at following attributes which can be used to predict the severity of accidents:

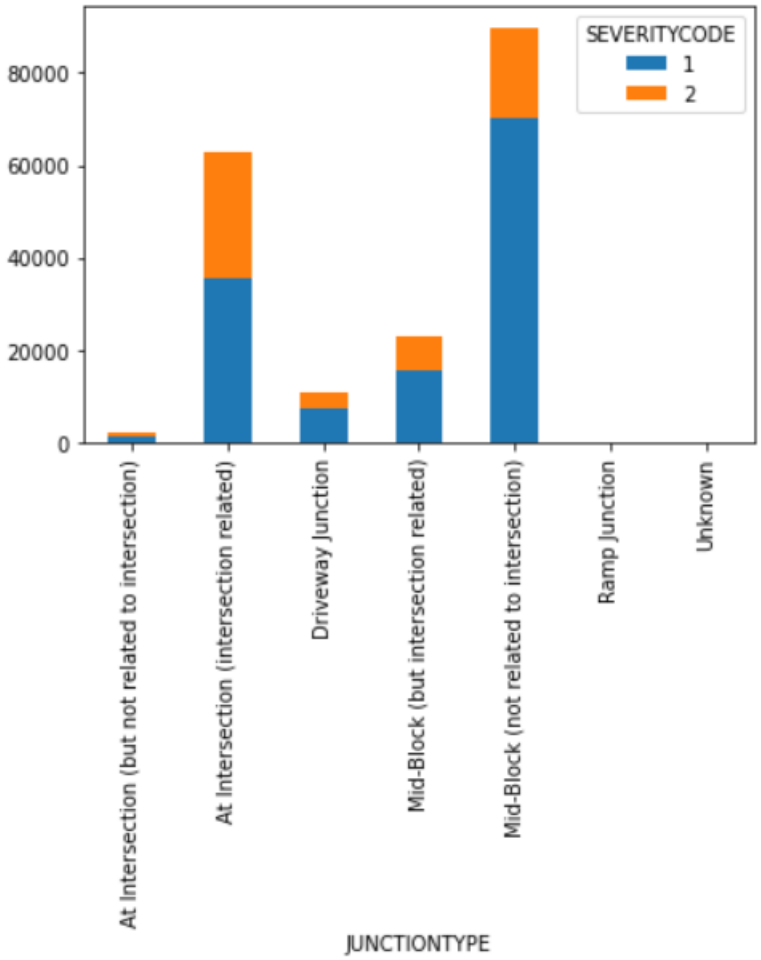
```
'ADDRTYPE', 'COLLISIONTYPE','PERSONCOUNT', 'VEHCOUNT', 'WEATHER', 'JUNCTIONTYPE', 'ROADCOND', 'LIGHTCOND'
```

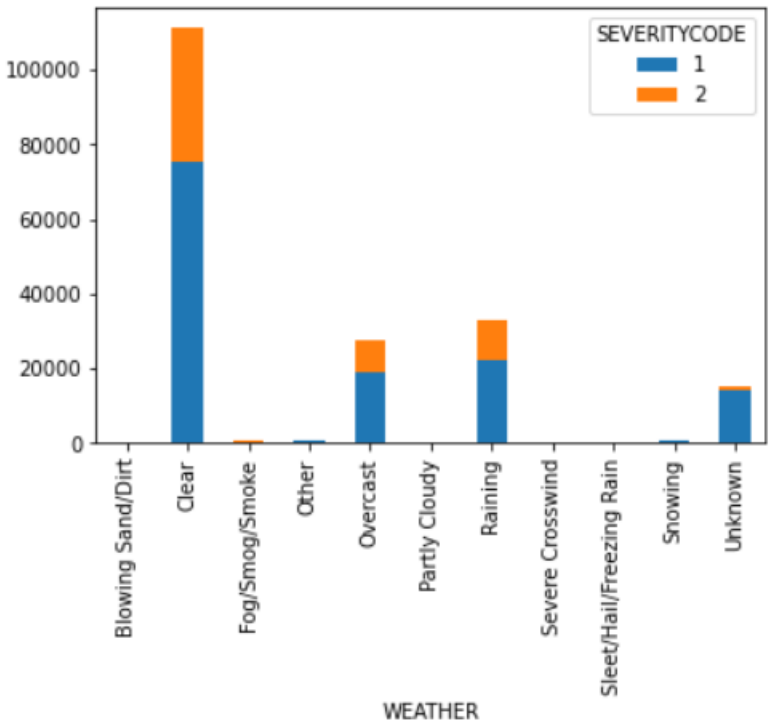


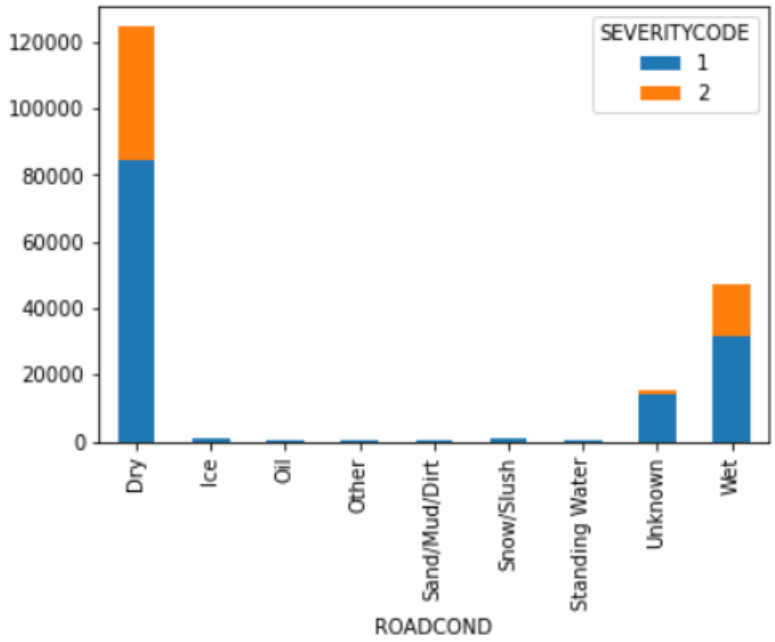

SEVERITYCODE
1
2
50000
40000
30000
20000
10000
0
Angles
Cycles
Head On
Left Turn
Other
Parked Car
Pedestrian
Rear Ended
Right Turn
Sideswipe
COLLISIONTYPE

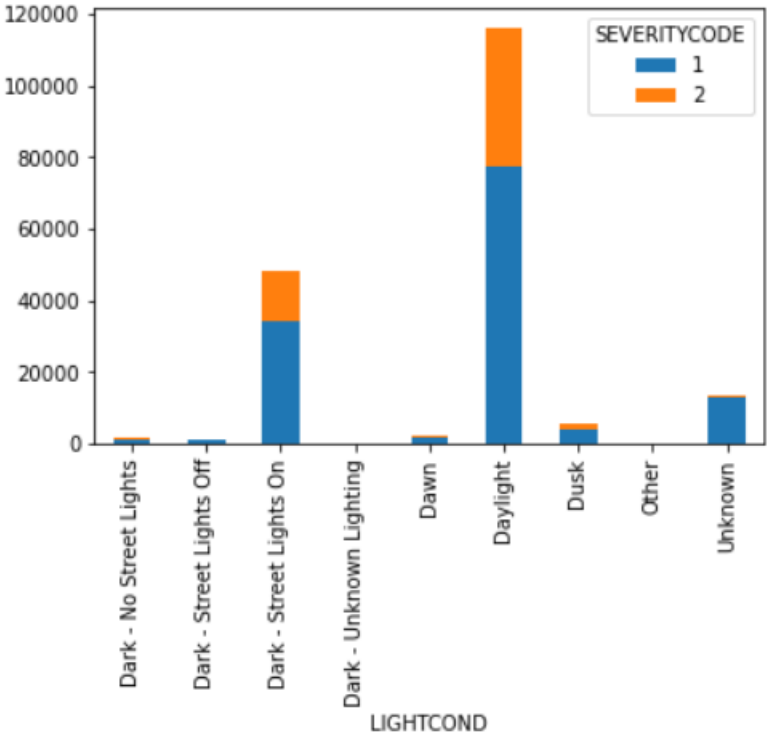

SEVERITYCODE
1
2
100000
80000
60000
40000
20000
PERSONCOUNT


SEVERITYCODE
1
2
175000
150000
125000
100000
75000
50000
25000
PEDCOUNT


SEVERITYCODE
1
2
80000
60000
40000
20000
0
At Intersection (but not related to intersection)
At Intersection (intersection related)
Driveway Junction
Mid-Block (but intersection related)
Mid-Block (not related to intersection)
Ramp Junction
Unknown
JUNCTIONTYPE


SEVERITYCODE
1
2
100000
80000
60000
40000
20000
0
Blowing Sand/Dirt
Clear
Fog/Smog/Smoke
Other
Overcast
Partly Cloudy
Raining
Severe Crosswind
Sleet/Hail/Freezing Rain
Snowing
Unknown
WEATHER





Data Cleaning

Once we understand the data, I need clean the data. I have done the following to clean the data:

1. Replacing Null value with Frequently occurring instances
2. Dropping rows for following attributes having null values:
    a. ADDRTYPE

b. JUNCTIONTYPE

Data Conversion

Data needs to be converted into numerical values. Values are assigned from 0 to 9 till we have different values.

Data Normalization, Splitting & Modelling

Data normalization was performed using sklearn. Post that data being split into Training & Test data in 80-20 ratio. In this case, since I have to predict the severity of accident, I have used classification method and kNN modeling is used.

Result

After model evolution, we for the below accuracy :

```
Train Accuracy:  0.72389652352554
Test Accuracy:  0.7018057069914635


KNN F1-score: 0.70
```