ACCIDENT PREDICTION

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.

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.

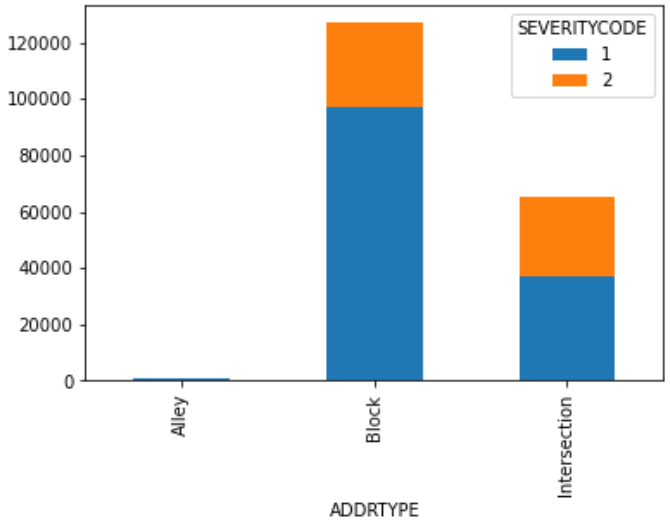
DATA DESCRIPTION

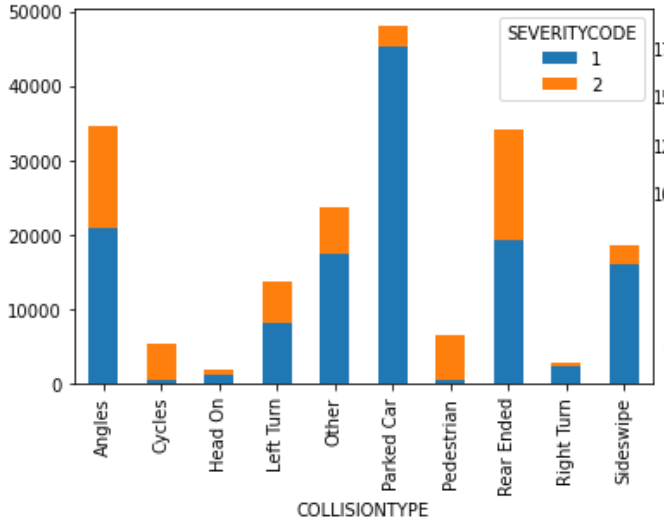
- Data downloaded from Coursera site in CSV format.
- Contains the 37 attributes of each accident.
- Most important is Severity code which we have to predict. Severity code is either 1 or 2.
- Other attributes which we can use are lighting conditions, road conditions, Weather conditions etc
- Above data will be used for both training & testing the model.
- Data is not cleaned as we see null values and some values are marked as “unknown” or “other”

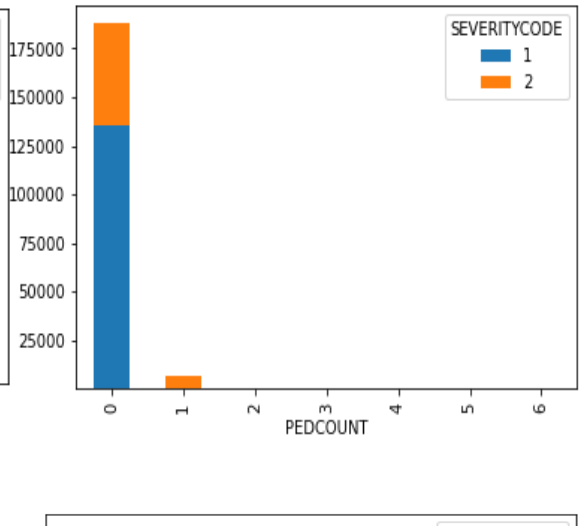
DATA ANALYSIS

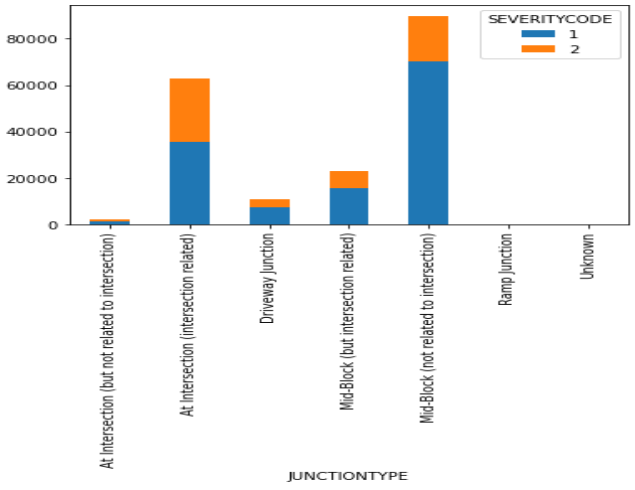
- After exploratory analysis, we arrived on following list of attributes which will be used for modelling
- ADDRTYPE’,
- 'COLLISIONTYPE’,
- 'PERSONCOUNT’,
- 'PEDCOUNT’,
- 'PEDCYLCOUNT’,
- 'VEHCOUNT’,
- 'WEATHER',
- 'JUNCTIONTYPE’,
- 'ROADCOND’,
- 'LIGHTCOND’,
- 'SEVERITYCODE'

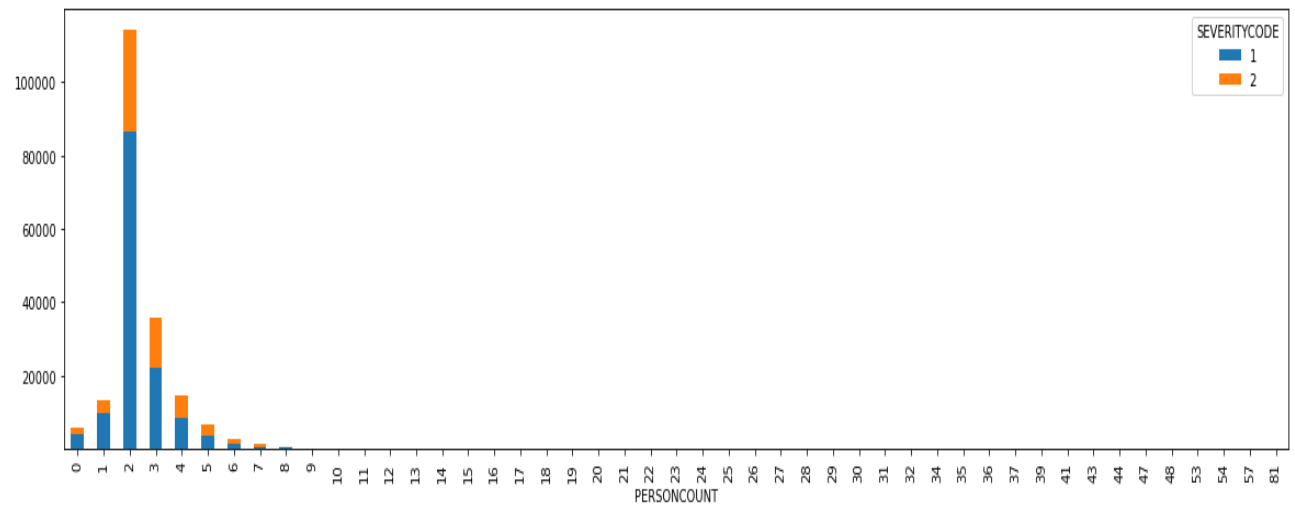
EXPLORATORY RESULTS

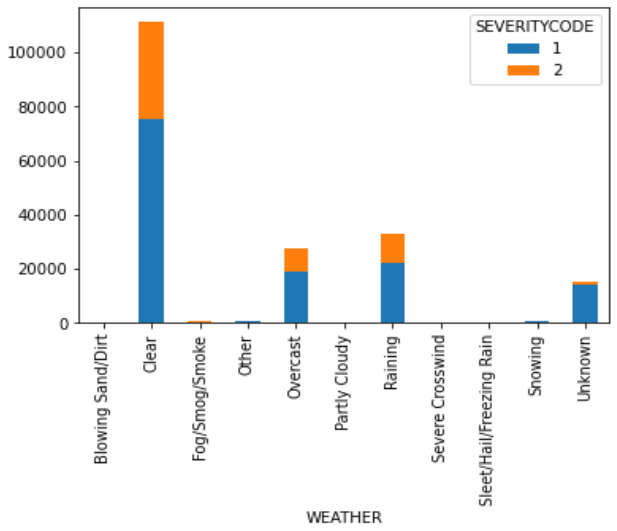


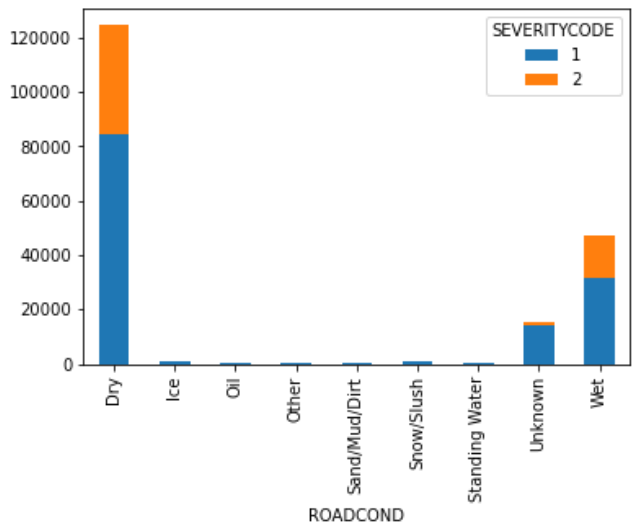












DATA CLEANING & NORMALIZATION

- Following aspects are taken into account for data cleaning
- Rows with attributes having null values, or values like “Other” or “Unknown”, are dropped
- Since we have categorical values, we have used two types of encoding:
  - Label encoding
  - One-Hot encoding
- Post encoding, we did the normalization of data before modelling

MODELLING & RESULTS

- Two types of modelling has been used:
- KNN
- Decision Tree
- Below is the results of these models

Modelling	F1 Score	Jaccard
KNN	0.69	0.72
Decision Tree	0.69	0.725

CONCLUSION

- Based on results we can say that both models have similar score and accuracy. However decision tree model is taking less duration to process data compared with KNN.
- So we will go for **Decision Tree modeling** and predict with more accuracy.
- We hope this will help traffic personals , government officials to reducing accidents and save lives.