CAR ACCIDENT SEVERITY PREDICTION

A Machine Learning Model

Presented by Laura. Q

Business Problem

- What are the factors that will influence the severity of a car accident? The weather? The number of people in the car? Or the location and road condition?
- If we can find the factors that are likely to cause a severe car accident from past real reports, we can try to lower to severity in the future. If we can predict the severity of a car accident given some known condition, we can make alerts to drivers and passengers when needed.
- This model will greatly reduce the loss of lives and properties from car collisions.

Data Description

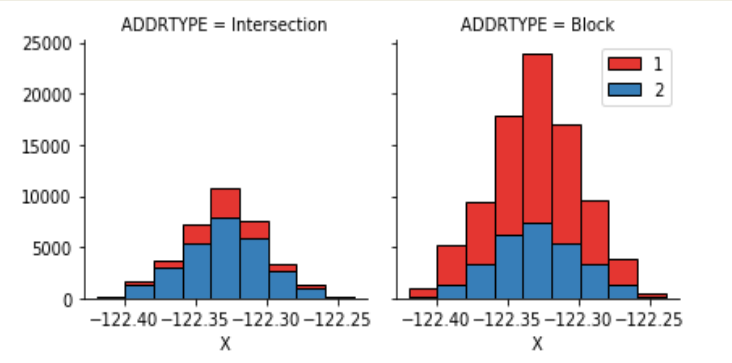
- The dataset includes all types of collisions happened in Seattle from 2004 to present recorded by Traffic Records of Seattle Government.
- There are 37 attributes in total but not all attributes will be used to build the model.
- The severity of a collision is recorded using a code from 0 to 3. Other attributes include location, collision type, date, time, weather, road condition, etc.
- Most attributes are strings but some are numbers.

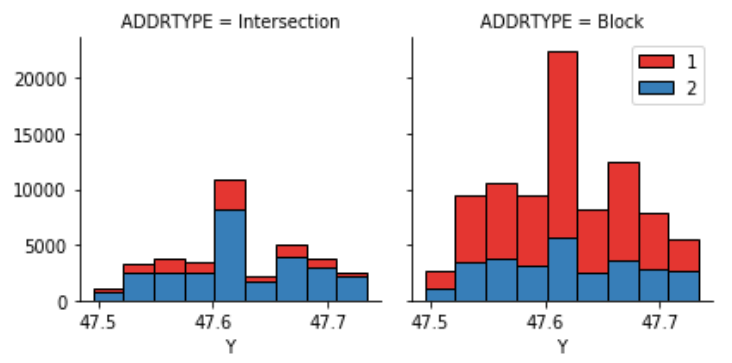
Methodology

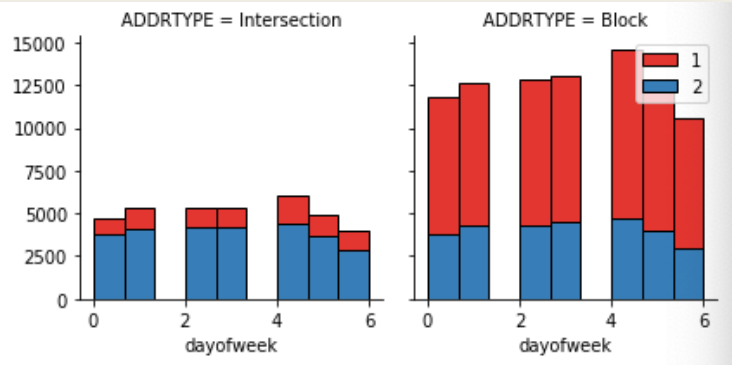
- 13 factors that are decided before an accident happen, that means, can be used for prediction have been chosen in the model.
- As for factors that are recorded after the accident are not considered here because we cannot know how many people will get injured in a collision or will there be a bicycle involved.
- For weather and road condition and collision type, after group by these factors I find that every attribute will result in different proportion of severity. The solution is to transfer them into dummy variables. It makes my final training set a bit wide with 37 features in total and above 190,000 observations.
- The machine learning method I choose to build a model is the K nearest neighbor.

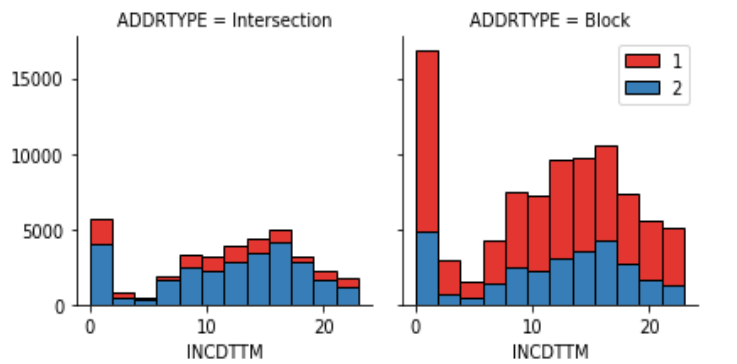
Results

- In some locations, severe accidents are more likely to happen,
- Collisions happened during the weekends and early mornings are less severe









Results

- whether drivers were speeding or whether pedestrians had obeyed the traffic rules will increase the probability of severe accident by a large amount.
- The difference of speeding is 8% and that of pedestrian not granting rule reaches 61% on average.
- If the collision is hitting a parked car, the probability of severe accident (code=2) will decrease by approximately 24% precent.

Future Research Direction

- The model reaches an accuracy of 0.7 but it can be improved from many other aspects. Future investigations could put into the pre-processing of data and exploration of other models.
- I dropped some NaN values and assumes some others are zero. In real cases I think we need to talk with the people who take these records or maintain the database.
- I am wondering if there are too many dummies that make the running speed of my code becomes a little bit slow. But it is hard to divide these dummies into groups because there is not enough correlation between the proportion of severity and the dummies.

Potential Application

- In this project I build a model to predict the severity of a car accident on some pre-decided conditions. It can be used in our transportation system for alert and reminder. Transportation department might be interested in it because they will know where and when to send warning messages to drivers. Other business that in transportation industries might want to utilize this model too. For example, it can be used in the auto-driving system and GPS guiding system.