Severity of Accident in Seattle City

A Coursera Capstone Project by
Nabeel Sharaf

1. Introduction

- Accidents are almost unavoidable in our society. City Planners must come up with innovative solutions to decrease the severity of them.
- This problem requires a scientific approach to be dealt with. As more accurate and comprehensive accident data are recorded, it gives the researches a better view in analysing these incidents and come up with a solution.

2. Data

- The Data used in this project is SDOT Traffic Management Division, Traffic Records Group. It has the information on all types of collisions provided by SPD and recorded by Traffic Records from 2004-2020, updated weekly.
- It contains a total of 37 attributes which will help in training the model and give a better prediction. Some of the entries in the data will be omitted to fine tune the model and get a better result.

3. Methodology

Data Analysis

Data Encoding

Feature Set and Normalization

Machine Learning Model

3.1 Data Analysis

- In order to identify the correlation of each attributes with the severity code, the below code is used

Dataframe.groupby([‘attribute’])[‘SEVERITYCODE’].value_counts()

- This will also help in eliminating columns which will hinder the accuracy of the machine learning algorithms.

3.2 Data Encoding

- It is important that we encode the datasets before starting the modelling. Most machine learning models only work with numerical data.
- Therefore, it is important that the other data types are converted to numerical data using the syntax:

Dataframe['column'].replace(to_replace=['A','B','C'],value=[0,1,2],inplace=True)

3.3 Feature Creation

- Now the feature set is prepared for the machine learning model to work on. Here the Feature set has 12 attributes

```
SEVERITYCODE      int64
ADDRTYPE        float64
JUNCTIONTYPE    float64
COLLISIONTYPE   float64
VEHCOUNT          int64
PEDCYLCOUNT       int64
PERSONCOUNT       int64
PEDCOUNT          int64
ROADCOND        float64
LIGHTCOND       float64
WEATHER         float64
HITPARKEDCAR      int64
dtype: object
```

3.4 Machine Learning Modeling

- The data is split into Train (80 %) and Test (20%) for evaluation purposes. The data is then run on K Nearest Neighbour and Decision Tree Machine Learning Models.

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=4)
```

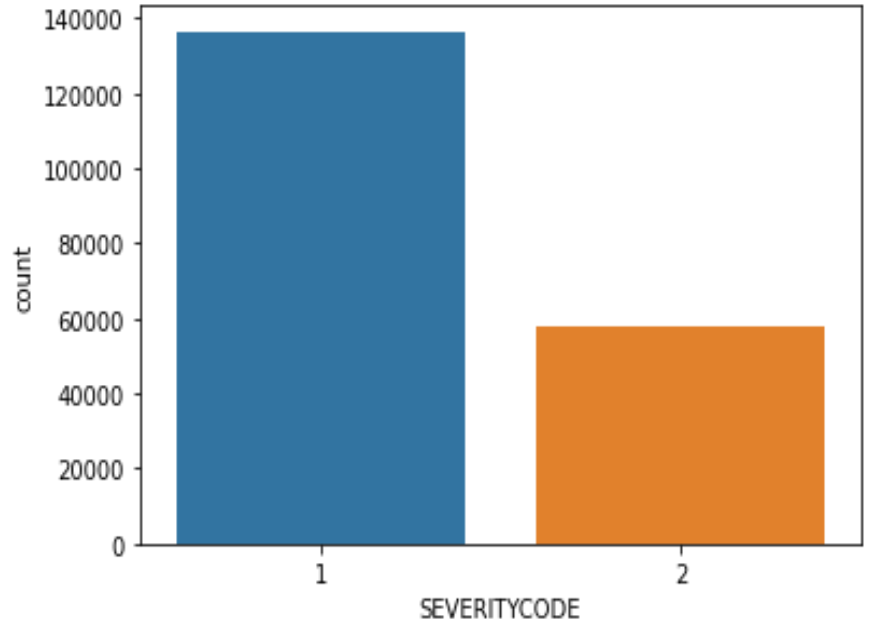
4. Result

The data is applied with K Nearest Neighbour and Decision Tree Machine Learning Models. The Jaccard Score anf F1-score was calculated to determine the accuracy of each model. For this project both the models used delivered similar results

Machine Learning Algorithm	Jaccard Score	F1-Score
K Nearest Neighbour	0.74	0.71
Decision Tree	0.74	0.71

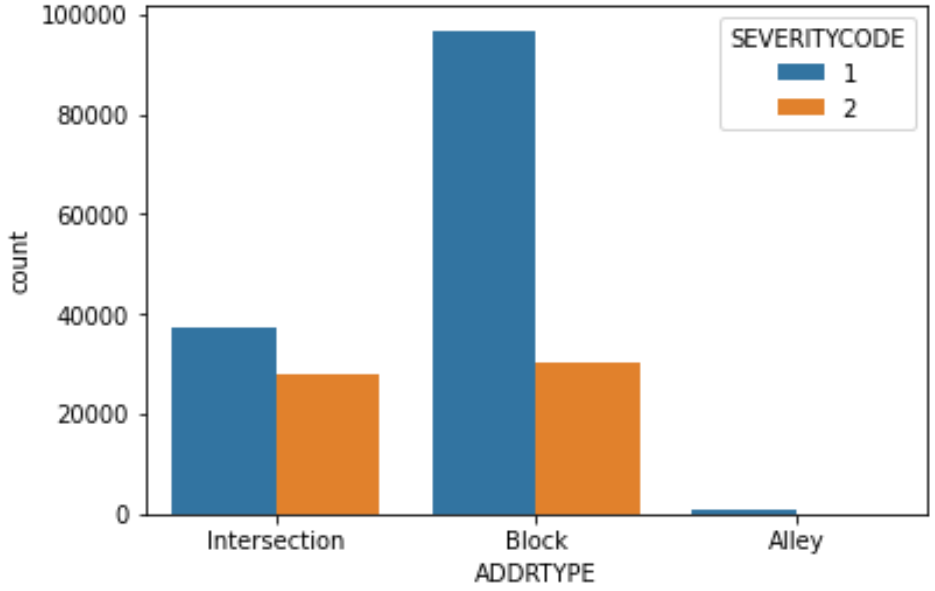
5. Discussion

- In this project the accident severity were analysed and a model to predict the severity of the accident was created.
- It is evident from the data that more accidents caused damage to the property than injury.



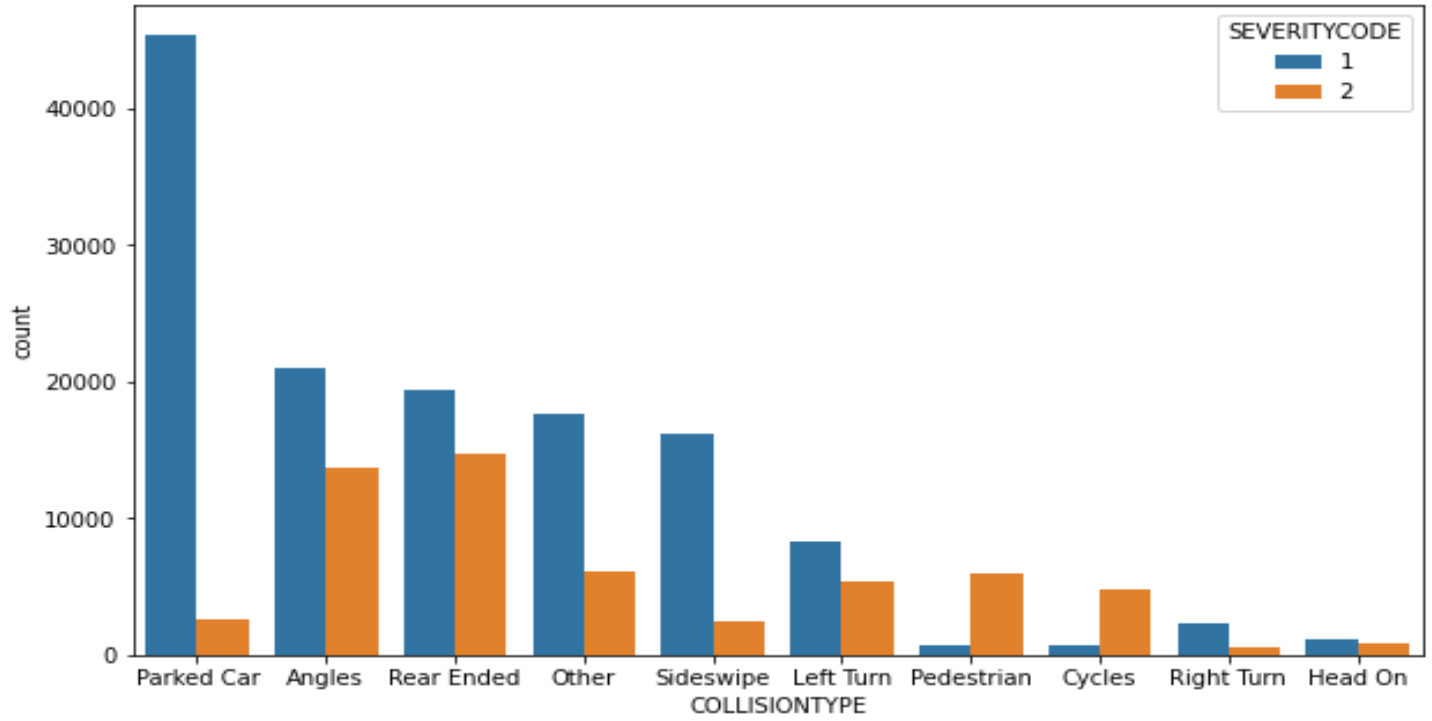
5. Discussion (Cont.)

- After analysing the data, it's understood that accidents happen more at Block Addresses than intersections or alleys. So it is advised to be more attentive at the addresses.



5. Discussion (Cont.)

- The most common type of collision was parked car collisions for property damage collisions, but collisions at an angle caused the most number of injuries.



Conclusion

- This project shows the advantages of using graphs and machine learning models to prevent or reduce road accident severity.
- A driver may look at these analytics to identify areas of possible accident causes and be prepared for such occurrences.
- The machine learning model may be used to predict the future outcomes and help the city management to create a plan to reduce the severity of such accidents.
- With more and more data's being available, machine learning model’s accuracy will be improved and a safer traffic is assured.