CAR ACCIDENT SEVERITY - SEATTLE

Chunduri Sumanth

Applied Data Science Capstone

Reducing Car Accident Severity in Seattle, Washington, USA

- The global epidemic of road crash fatalities and disabilities is gradually being recognized as a major public health concern.
- Approximately 1.35 million people die in road crashes each year, with the daily global death toll being about 3,700 people. An additional 20-50 million suffer non-fatal injuries every year, often resulting in long-term disabilities (https://www.asirt.org/safe-travel/road-safety-facts/).
- Accidents that occur can be slight, fatal and serious. As such, a good mitigation tactic is to be able to reduce the risk of severity before accidents occur. In an effort to reduce the frequency of car collisions in a community, an algorithm can be developed to predict the severity of an accident given the weather, road and visibility conditions.

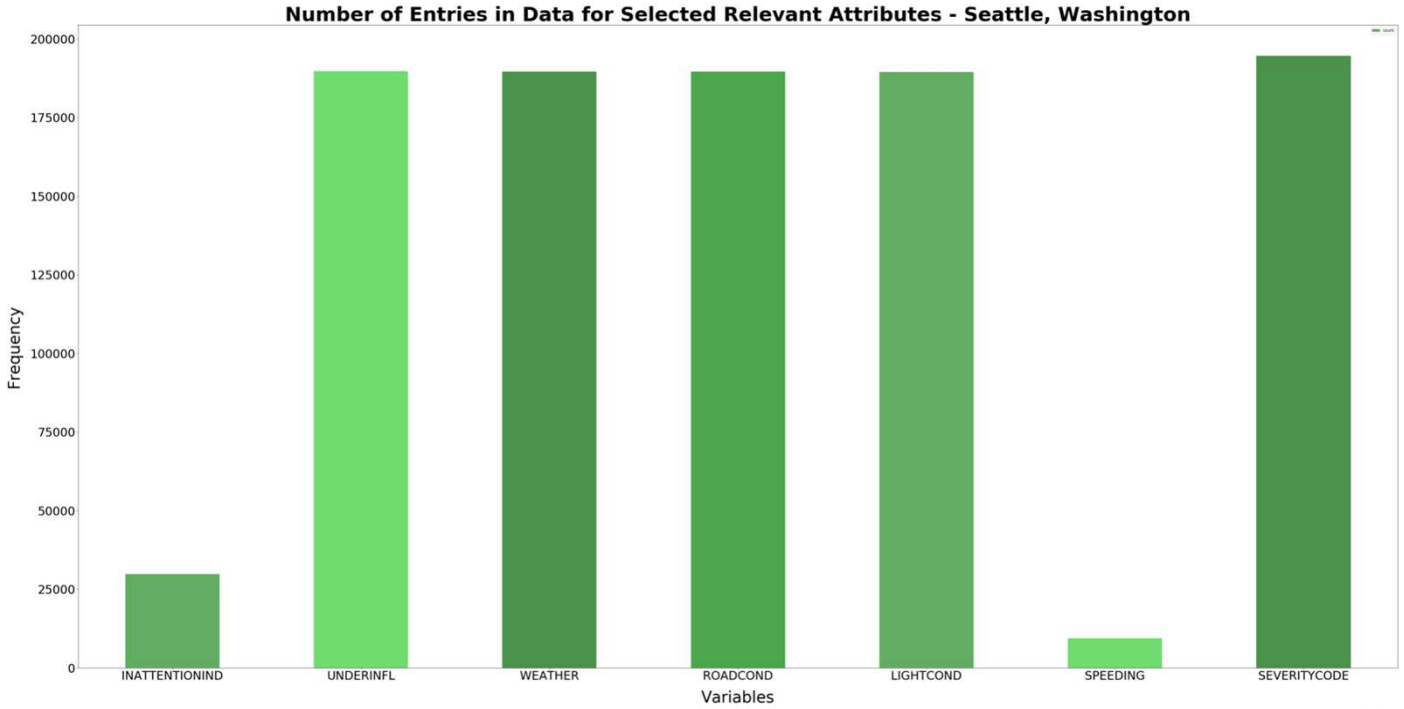
Data Acquisition and Cleaning

- The dataset has information gathered on the road traffic accidents of Seattle City.
- The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present. The data consists of 37 independent variables and 194,673 rows.
- The dependent variable, “SEVERITYCODE”, contains numbers that correspond to different levels of severity caused by an accident from 0 to 4.
- The dataset can be found here: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv.

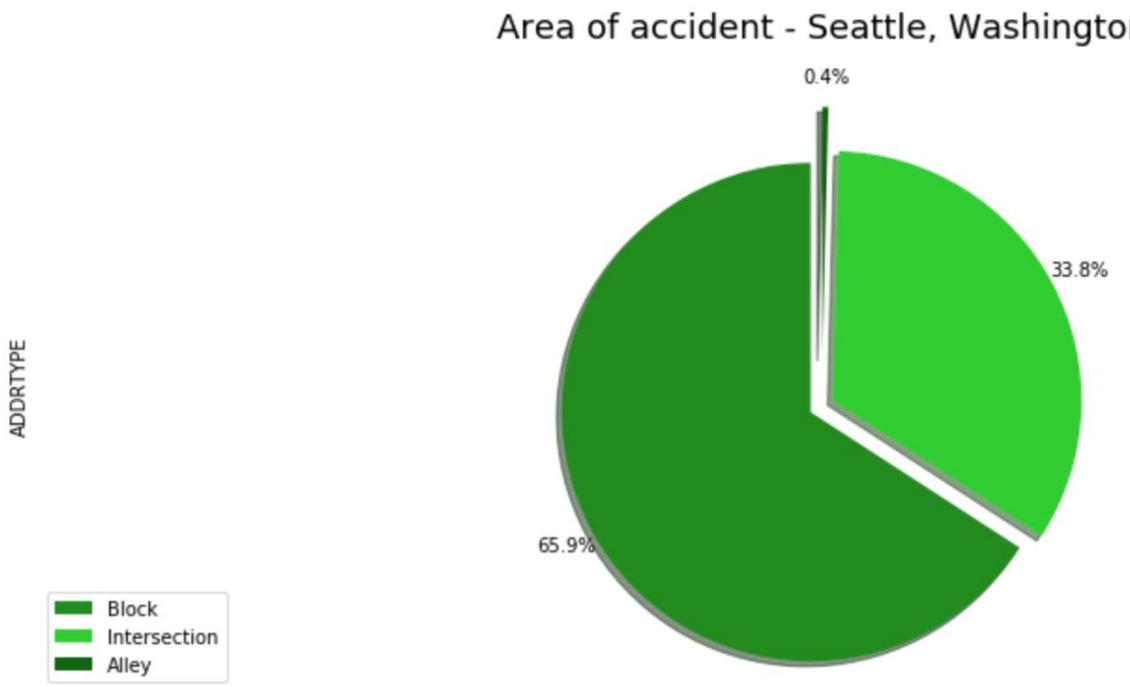
Feature Selection

- The following features were selected for this machine learning project:
  - WEATHER - Weather condition during time of collision (Overcast/Rain/Clear)
  - ROADCOND - Road condition during the collision (Wet/Dry)
  - LIGHTCOND - Light conditions during the collision (Lights On/Dark with light on)
  - INATTENTIONIND - Whether or not the driver was inattentive (Y/N)
  - UNDERINFL - Whether or not the driver was under the influence (Y/N)
  - SPEEDING - Whether the car was above the speed limit at the time of collision (Y/N)

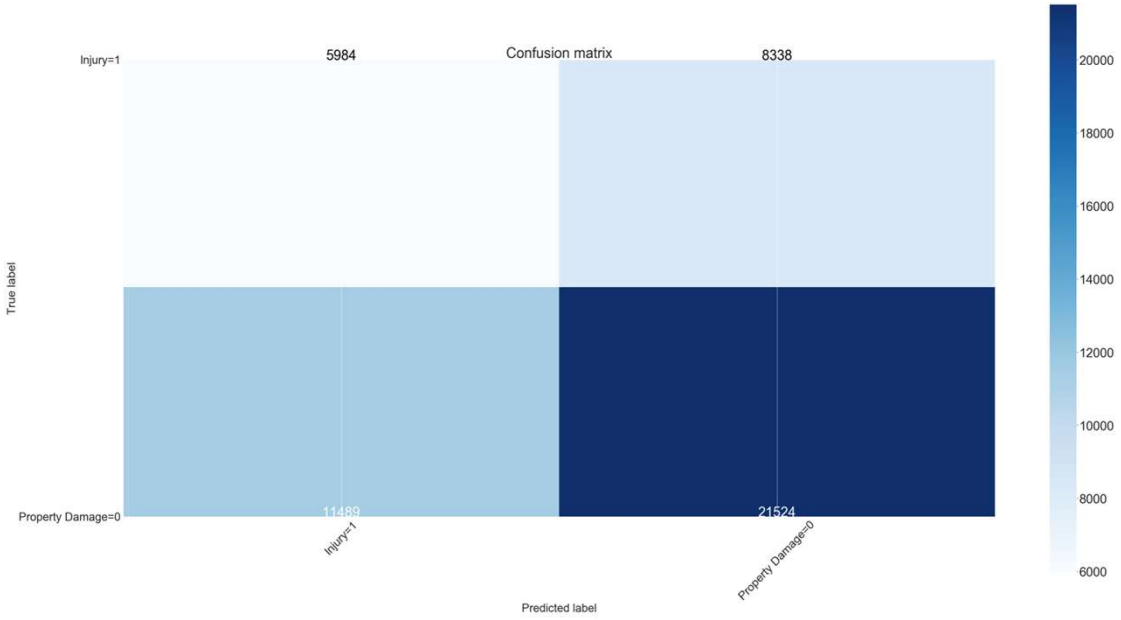
Exploratory Data Analysis



Exploratory Data Analysis

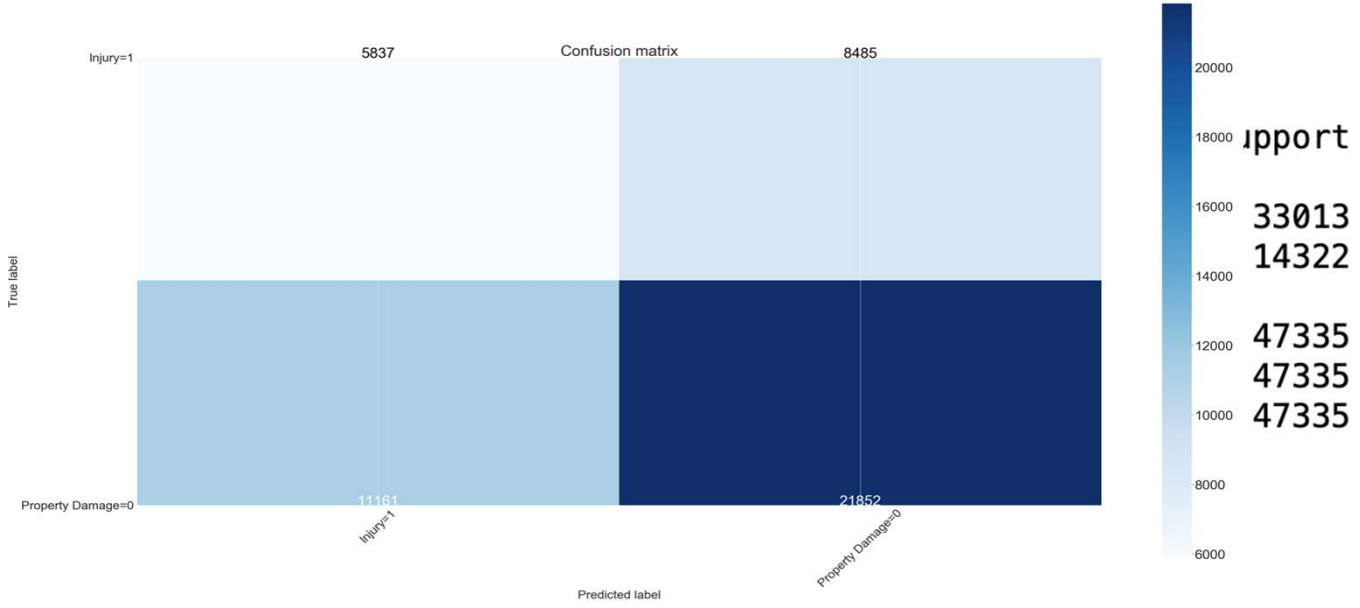


Decision Tree Algorithm



	precision	recall	f1-score	support
0	0.65	0.72	0.68	29862
1	0.42	0.34	0.38	17473
accuracy			0.58	47335
macro avg	0.53	0.53	0.53	47335
weighted avg	0.57	0.58	0.57	47335

Logistic Regression



ıpport

33013
14322

47335
47335
47335

k-Nearest Neighbour (kNN)

```
KNNf1score
```

```
0.555808952766952
```


```
Best K is : 10 | Cross validation Accuracy : 0.690461899717634
```

Conclusion and Model Selection

- When comparing the f1 scores, accuracy, confusion matrixes, and classification reports of the three different algorithms, we are able to better determine which algorithms have performed better in achieving the goal of the project.
- Evident by the various scores, it is clear that k-Nearest Neighbour algorithm is the most suitable for this project.

	Accuracy	F1 Score
Decision Tree	0.58	0.53 (0: 0.68; 1: 0.38)
Logistic Regression	0.58	0.53 (0: 0.69; 1: 0.37)
kNN	0.69	0.56

Recommendations and Insights

- The Seattle municipal government can assess areas in the city with high accident rates and focus on implementing preventative measures to try to mitigate the accidents. For example, development projects could be launched to target improved light conditions in high-risk areas.
- Since almost all of the accidents in Seattle happen in either a block or an intersection, the municipal government can develop initiatives to mitigate unfavourable conditions in those areas. Those can include installing better safety signs to notify drivers of hazards, as well as increase investment in lightening and road conditions in high-risk areas.
- Additionally, individual drivers can use this data to be cognizant of areas and conditions that have a higher risk of getting into a car accident. Specific importance can be given to mitigate conditions and areas that cause severe car accidents.