Predicting the severity of Road accident based on Traffic Incident data

Parag S.

The Prediction Made on Time could be Beneficial to the drivers and Help save life and Time

- According to World Health Organization, every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash.
- Seattle has the 8th worst traffic congestion of all American cities, and is 10th among all North American cities according to Inrix
- Seattle recorded the highest number of car accidents in the state in 2018, at 14,508 in Washington.
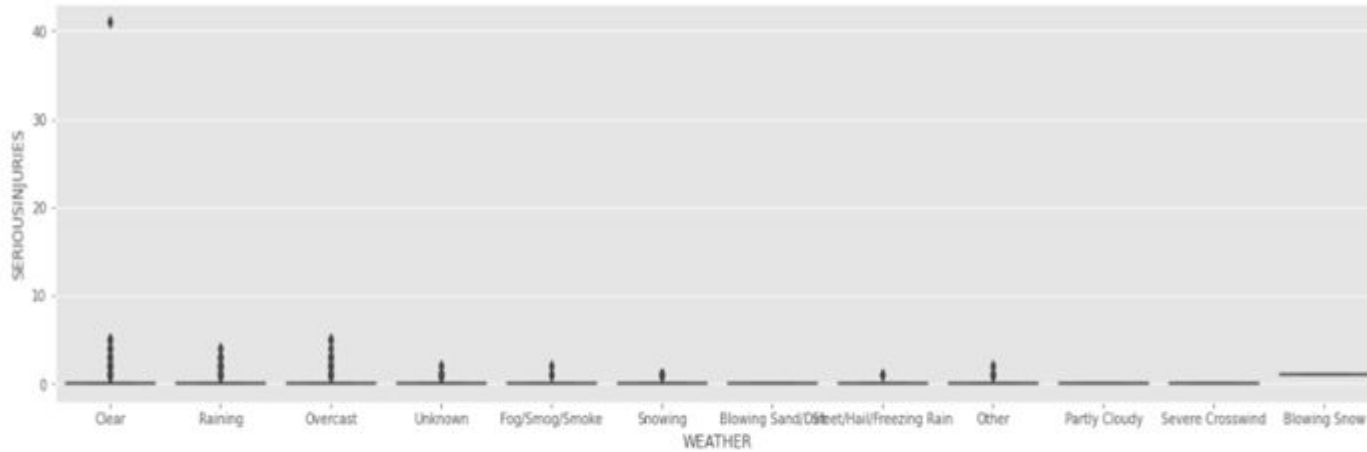
Data acquisition and cleaning

- Data Acquired from City of Seattle at the open data platform (https://data.seattle.gov)
- The data was heavily biased so it needed to be balanced.
- One Hot encoding was done on data.
- Final test and train data-

```
Train set: (77286, 30) (77286,)
Test set: (33123, 30) (33123,)
```

The data had uneven distribution among features
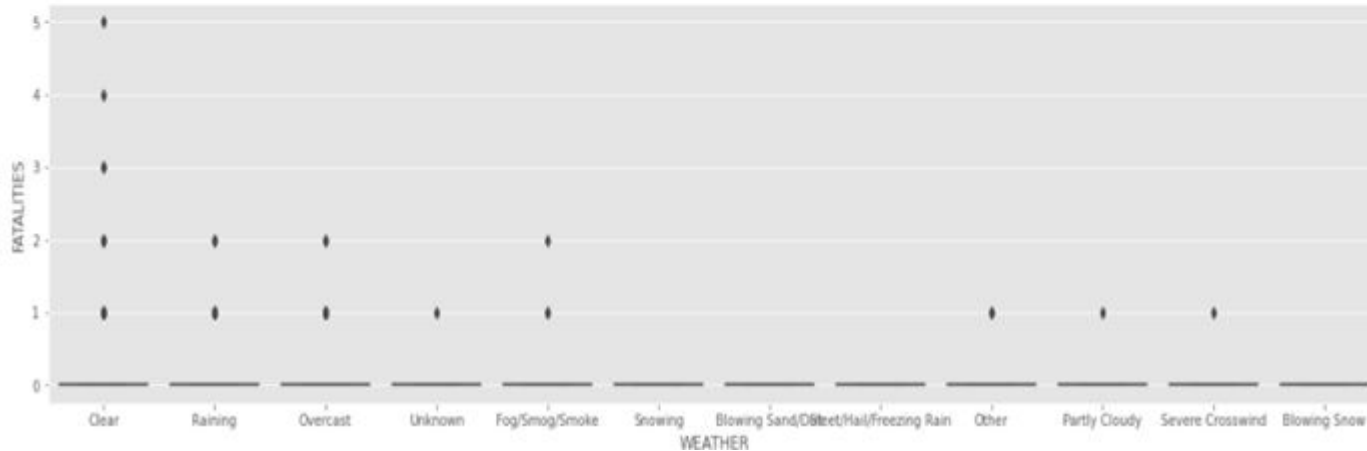
All the features were analyzed

- The Features were Analyzed and Unknowns were dropped.
- Along With this the Feature type which had very less number of observations were dropped.
- The data was cleaned unavailable values were dropped.

Classification Models Used

- KNN
- Decision Tree
- Support Vector Machine
- Logistic Regression

Model Scores Based on the Best parameters Selected

Algorithm	Jaccard	F1-score	LogLoss
Logistic Regression	0.644416	0.644726	0.608068
Decison Tree	0.651300	0.649031	NA
KNN	0.626936	0.626915	NA
SVM	0.644114	0.642526	NA

Conclusion and Future Direction

- Based on the performance matrices we choose the Decision Tree as the best classifier for our model.
- The created model's efficiency can be increased by concatenating other features to our feature set. From other Data Sources.
- The built model was able to predict the test data with 65% accuracy