Applied Data Science Capstone Project

Car accident severity

1. Introduction/Business Problem

- The project aims to predict how severity of accidents can be reduced based on a few factors such as weather, road, light conditions etc.
- This project may be beneficial to the Public Development Authority of Seattle and the car drivers.

2. Data Understanding

- Severity of accidents in Seattle city is the Data set.
- Predictor or target variable will be 'SEVERITYCODE‘.
- Attributes used to weigh the severity of an accident are 'COLLISIONTYPE', 'WEATHER', 'ROADCOND','LIGHTCOND','JUNCTIONTYPE', 'LOCATION', 'PERSONCOUNT' and 'VEHCOUNT'.

3. Data Preparation

- The Dataset was corrected for missing values.
- The dataset was found to be imbalanced.
- Using Undersampling some of the observations from the majority class were deleted in order to match the numbers with the minority class.
- Label Encoding was used to convert the categorical data into numeric form so as to convert it into the machine-readable form.

4. Modelling

- The machine learning models used are:

i. SVM
ii. Decision Tree
iii. K-Nearest Neighbour.

5. Result

- Classification report shows that the Decision Tree model has 70% accuracy.

	precision	recall	f1-score	support
1	0.75	0.59	0.66	16806
2	0.67	0.81	0.73	17152
accuracy			0.70	33958
macro avg	0.71	0.70	0.70	33958
weighted avg	0.71	0.70	0.70	33958

- Classification report shows that the KNN model has 67% accuracy.

	precision	recall	f1-score	support
1	0.70	0.59	0.64	11315
2	0.65	0.74	0.69	11324
accuracy			0.67	22639
macro avg	0.67	0.67	0.66	22639
weighted avg	0.67	0.67	0.66	22639

- Classification report shows that the SVM model has 70% accuracy.

	precision	recall	f1-score	support
1	0.74	0.60	0.66	16806
2	0.67	0.80	0.73	17152
accuracy			0.70	33958
macro avg	0.71	0.70	0.70	33958
weighted avg	0.71	0.70	0.70	33958

- **Average f1-score**

f1-score of Decision tree and SVM are almost the same (0.66) where as that of KNN is 0.63.

From these results we can assume that all the three f1-scores fairly good in predicting Property Damage and Injury.

- **Precision**

The precision level to predict the “Property damage Collision” of Decision tree is highest (0.75) and that of KNN is lowest (0.70). However, the precision level to predict the “Injury Collision” of all the 3 models is almost same.

From these results we can assume that Decision tree is fairly good in predicting Property Damage and Injury.

6. Conclusion

- The project may be beneficial to the Public Development Authority of Seattle and the car drivers as it may be useful in preventing future accidents in the city.