CAR ACCIDENT SEVERITY REPORT:SEATTLE,WASHINGTON

BY ABDUL SABOOR

DATE: 21/10/2020

Exploratory Data Analysis:Why it is important to reduce car accidents?

► Seattle is the largest city in the state of Washington , and is a hub to large two tech giants Microsoft and Amazon.

► Seattle accounts for nearly 3.4 million population, car accidents has become a major issue a lot these days due to increased car population

► Nearly almost 1.25 million people die in road crashes each year. Car accidents are one of the leading causes of death.

► It took a toll of 518 billion USD on US government. According to Seattle Times, the city's goal is to achieve zero fatalities and serious injuries by 2030.

Data Source:
The data has been provided by SPD(Seattle Police Department) and recorded by Traffic Records Department. The data set has total observations(rows) of 194,673.The severity code is as follows:

SEVERITY CODE	DESCRIPTION
3	Fatality
2b	Serious Injury
2	Injury
1	Prop Damage
0	Unknown

Methodology:

In this project we have used most significant feature variables like "WEATHER","ROADCOND" and "LIGHTCOND" to predict our target variable or outcome which is "SEVERITYCODE" in this case .Lets look at the table to get further understanding:

FEATURE VARIABLES	DESCRIPTION
WEATHER	Weather condition during the time of collision(wet,dry,clear)
ROADCOND	Road condition during the collision(Wet or Dry)
LIGHTCOND	Conditions of light during collision(bright or dark)

Modeling and Evaluation

▶ Following machine learning models are applied Logistic Regression, K-Nearest Neighbor, Decision Tree.

▶ The reason we are not using SVM Support Vector Machine Model is because they are inaccurate for large data sets, Hence SVM works best for the data which filled with text and images.

▶ Furthermore after preprocessing and scaling the data we applied machine learning models,

Results

Machine Learning Model	Jaccard-Index	F1 Score	Accuracy
KNN	0.237	0.512	0.546
Decision Tree	0.285	0.543	0.564
Logistic Regression	0.277	0.515	0.529

Conclusion

- Based on the results we can see Decision Tree is the best machine learning model.However these models could have done better if we have more balanced dataset for target variable,

- Factors like precautionary measures when driving and etc.