Predicting Traffic Accident Severity

Business Understanding:

In an effort to reduce the frequency of car collisions in a community, an algorithim must be developed to predict the severity of an accident given the road, current weather and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be careful.

Data understanding:

The dataset used for this project is based on car accidents from the year 2004 to 2020. This data is regarding the severity of each car accidents along with the time and conditions under which each accident occurred. .

About Feature Selection，the predictor or target variable will be 'SEVERITYCODE' because it is used measure the severity of an accident from 0 to 5 within the dataset. Attributes used to weigh the severity of an accident are 'WEATHER', 'ROADCOND' and 'LIGHTCOND'.

Data preparation was performed multiple times and it includes balancing the labeled data, transformation, filling missing data, and cleaning the dataset.

The target variables between Physical Injury and Property Damage can be seen that the dataset is supervised but an unbalanced dataset where the distribution of the target variable is in almost 1:2 ratio in favor of property damage. It is very important to have a balanced dataset when using machine learning algorithms.

Modeling:

Machine Learning Models chosen

- **Logistic Regression**
- **KNN**
- **Naive Bayes**
- **Decision Tree Analysis:**
- **Random Forest Classifer**
- **SVM**

Evaluation:

Both F1 score and Accuracy of each model are not that good.

Conclusion：

Anyway, after assessing the data and the output of the Machine Learning models, a few recommendations can be made for the stakeholders. The developmental body for city can assess how much of these accidents have occurred in a place where road or light conditions were not ideal for that specific area and could launch development projects for those areas where most severe accidents take place in order to minimize the effects of these two factors. Whereas,

the car drivers could also use this data to assess when to take extra precautions on the road under the given circumstances of light condition, road condition and weather, in order to avoid a severe accident.