

Business Problem

The Seattle government is going to prevent avoidable car accidents by employing methods that alert drivers to remind them to be more careful in critical situations.

Besides the common reasons that cause accident, weather, visibility, or road conditions are the major uncontrollable factors that can be prevented by revealing hidden patterns in the data and announcing warning to the local government and drivers on the targeted roads.

The target audience of the project is local Seattle government, drivers, rescue groups. The model and its results are going to provide some advice for the target audience to make insightful decisions for reducing the number of accidents and injuries for the city.

Data

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.

The data consists:

- 37 independent variables
- 194,673 rows.
- The dependent variable **SEVERITYCODE**, contains numbers that correspond to different levels of severity caused by an accident from 0 to 4.

Severity codes are as follows:

- 0 → Unknown
- 1 → Chance or Property Damage
- 2 → Chance of Injury
- 2b → Chance of Serious Injury
- 3 → Chance of Fatality

Data Preprocessing

The dataset in the original form is not ready for data analysis. In order to prepare the data

- I use **Resample** from **sklearn.utils** package to balance data.
- Most of the features are of object data types that need to be converted into numerical data types.
- Drop Unimportant Columns.

I have decided to focus on only four features:

- Severity
- Weather Conditions
- Road Conditions
- Light Conditions

Methodology

For implementing:

- Github as a repository and running Jupyter Notebook.
- Build Machine Learning models via Python libraries such as pandas, numpy and sklearn.

I have used KNN Model, Decision Tree Model and Logistic Regression Model to predict the severity of accidents through weather, road and light conditions. Then calculate the accuracy of each model to check which model is better to apply it to prevent accidents.

Result

Out[32]:

	Algorithm	Jaccard	F1-Score	Logloss
0	KNN	0.548220	0.514793	NA
1	DecisionTree	0.556841	0.468595	NA
2	LogisticRegression	0.506974	0.362167	0.693147

Conclusion

Referring to our dataset and after applying Machine Learning Models Accuracy so we can conclude that weather, road and light conditions have an impact on severity of accident if it's property damage or injury.