Predicting Car Accident Severity

-Ranga R

Predicting car accident severity is important to insurance companies

- Generally, insurance companies rely on claim adjusters to assess the severity of car accidents, so as to set aside appropriate amounts for claim reserves.
- By using a model to predict the severity of car accidents, insurers will be able to automate the process of setting up claim reserves and reduce reliance of claim adjusters' opinions.

Data Acquisition and Cleaning

- The dataset has been provided for Seattle Police Department and was recorded by Traffic Records.
- In total, there are 194673 rows and 36 features in the dataset.
- Duplicate features and features that do not provide sufficient information have been removed from dataset.
-

Methodology

- I have built several types of machine learning models to predict the claim severity.
  - K nearest neighbours
  - Decision tree
  - Logistics Regression
-

K Nearest Neighbors

- I had split the dataset into a training dataset and test dataset, where 80% of the original dataset is used for training while 20% is used for testing.
- I have then applied an algorithm k = 4 to build the k nearest neighbours model.
- The accuracy of the training and test models are as below:
  - Train set accuracy: 74.0%
  - Test set accuracy: 73.7%

Decision Tree

- Similar to the k nearest neighbours model, I had split the dataset into a training dataset and test dataset, where 80% of the original dataset is used for training while 20% is used for testing.
- I have then applied a maximum depth of 4 to build the k nearest neighbours model.
- The accuracy of the test model is 75.3%.

Logistic Regression

- Similar to the aforementioned models, I had split the dataset into a training dataset and test dataset, where 80% of the original dataset is used for training while 20% is used for testing.
- The accuracy of the test model is 75.5%.

Results

- Based on the accuracy evaluations performed for each of the models mentioned previously, logistic regression model has the highest accuracy.
- Therefore, it would be appropriate to use a logistic regression model to predict car accident severity in Seattle.