

Predictions for At Fault Drivers

Overview

Using Logistic Modeling and DecisionTreeRegression to predict the possibility of a driver being at fault for car crashes in the Chicago, IL area. The Data was provided by the Chicago, IL police department.

The Data

The Data used is the data provided by the Chicago data portal,PTraffic Chrashes-Crashes. The data is provided by the Chicago police department and includes weather condition, injuries, and roadway condition. The Data originally had 49 columns but only 23 columns were ultimately used.

Business Problem

Geico would like a way to predict the car crash cases that they should presue an investigation with. They would like to avoid allocating funds to a unnecessary investigations.

The Predictive Models

Initially, there was an attempt to check the class balance of the target data. The target data includes has two values, 0 for outside circumstances and 1 for Driver's fault. The class balance check found that the balance is very close to even with a slight bias to false positive. The balance was not adjusted because the sloght balance could benifit Gieco with finding at fauld Drivers.

![[image]](https://user-images.githubusercontent.com/62761238/119340133-16c7b080-bc60-11eb-82f7-923718c804c4.png)

The first 2 models were logistical models using different aspects of the data. The inital model only incorporated the numerical features from the DataFrame. The accuracy score was about 57%. In an attempt to increase the predictivity the categorical features were added in the second model. This addition did not help the model and dropped the accuracy model to 54%. Finally, a DecisionTreeRegression model was used as an attempt to change the predictive model and increaset the accuracy score. The score increase from the second logistical model but was still 1% less then the original