Peer-graded Assignment: Capstone Project - Car Accident Severity

Sreeja Matturu

September 27, 2020

Data

The data set is a CSV file format that consists of different road accidents that occurred in Seattle. The data set was acquired from Kaggle.com where it contains 37 different types of attributes and more than 190k rows of data describing each accident at each time. This data set contains outliers which are represented as NaN and thus represented as an unbalanced dataset. It includes different types of attributes like injuries, pedCounts, vehicle counts, and fatalities. Features that I want to use from this data set include categorical variables like type of collision, collision codes, whether or not the pedestrian was granted right of way, road, and weather conditions. Primarily, the entire data set will be considered as a feature set, then we can remove the null values to apply different machine learning techniques to predict the severity of a car accident as an outcome. Machine learning includes different classification and regression techniques like KNN, Decision Tree, SVM, and Logistic Regression. Applying these techniques, we can find the relation between the attributes and predict the severity of the car accident as well as the potentiality of a car accident as well.