Santiago Andrés Granda Bravo

09/30/2020

Applied Data Science Capstone

Capstone Project – Car accident severity Dataset (Week 1).

The dataset that will be analyzed for this project is the collisions dataset from Seattle from 2004 to present, shared by the instructors. The dataset contains information of the type of collision, location, weather, the severity of the collision, and many other attributes that occurred, which are reported and described on the dataset. The dataset contains approximately 200 thousand of samples with 37 features or columns. From the dataset, the "SEVERITYCODE" feature is the dependent variable. The values that can take the target feature are 1 or 2, where 1 indicates low severity or 2 high severity after a crash. All the other features must be analyzed to observe the correlation between the independent variables with the target variable. Examples of the independent variables are: "LOCATION" is a string type which indicates the general location of the collision such as: '5TH AVE NE AND NE 103RD ST', "VEHCOUNT" is a numeric double type variable that indicates the number of vehicles involved in the collision such as: '2', "WEATHER" is a string type variable and indicates the weather conditions during the time of collision such as: 'Raining', etc.

To develop a model with the dataset, first, it must be cleansed and preprocessed to perform the classification algorithm. Applying this method, the dataset is going to have the necessary attributes and features to obtain a good an acceptable solution.

Bibliography:

https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf

https://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0?selectedAttribute=SEVERITYCODE