Capstone Project / Business Understanding

Project Objective

This project will use collision severity data from Seattle in order to better understand what influences the severity of collisions.

In particular, the data will be used to understand what influences data *at intersections*.

Target Audience

The target audience for this project is city council, city planners, and traffic planners and engineers. The goal is for the model outputs to be used in influencing future intersection design to make for safer intersections. Additionally, the model could be used by that audience to address particularly dangerous existing intersections.

The Machine Learning Problem

The goal of the associated machine learning problem is to take the available data, clean and select atributes, and generate a predictive model for collision severity.

The model will be based on multi-class classification.

Since the goal is to identify the most important factors, the model will need to have a way to assess importance of specific attributes to the final classification.

The available data was retrieved here: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

The detailed description and metadata of the dataset can be found here: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf

Preliminary Plan

1. Assess the data

2. Identify the key attributes most strongly influencing collision severity

2a. Clean and transform as necessary

3. Build several machine learning models and assess for accuracy in predicting collision severity

4. Use attribute importance within the most accurate model to prioritize factors that most strongly influence collision severity. As collision severity is a multi-class classification problem, this will potentially limit the choices of algorithms.

5. Make preliminary investigative recommendations on possible intersection modifications (current and future)