Road Collision Severity in Junctions

Akil Sahib

September 7, 2020

1. Introduction

1.1 Background
Road transportation is a major form of transportation in modern society and road traffic accidents can cause a considerable economic loss to individuals and their families. These losses arise from cost of treatment to cost of alternative transportation to work etc. Road safety is considered one of the most important problems in road transportation. Accidents that occur in junctions can be slight, fatal and serious hence it is very important to be able to reduce the risk of severity before accidents occur.

1.2 Problem
Road traffic is prone to changes and can not very often be predicated or planned for accurate. But reducing the level of severity also means shorter amount of time in delaying the traffic. The aim for this project is to build a supervised classification model that can predict the severity of accidents in junctions around different types of addresses, collision, weather conditions and light conditions.

1.3 Audience:
The audience of this project are the local authorities and the police. For instance, this project can help lead local authorities discover the address type where severity of a road accident is serious due to lack of road lighting or warning signs for easy loss of control or sign ahead nor very clear.

2. Data

2.1 Data Source
The data used for this study is given by the Applied Data Science Capstone course on Coursera.org via the following link https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv. The dataset has information gathered on the road traffic accidents of Seattle City. Python packages will be used to conduct this study. The dataset will be cleaned according to the requirements of this project.

2.2. Data Cleaning
The data and the metadata will be first observed to allow for better understanding of what the attributes mean and their elements. Firstly, the data is read into pandas csv dataframe and displayed to be observed. We will also use shape and columns functions to obtain information on the shape of the dataframe and column attributes. For the interest of this project, the features/attributes interested in are:

- SEVERITYCODE: the level of severity - 1 = Property Damage Only, 2 = Injury Collision
- ADDRTYPE: block or intersection junction
- JUNCTIONTYPE: type of junction
- SDOT_COLDESC: description of the collision between parties involved
- WEATHER: weather condition
- LIGHTCOND: road light condition

The initial dataset consists of 38 columns (features/attributes) and 194673 rows. Using pandas drop function we drop all columns expect the desired ones. Columns with the same information but in either categorical or numerical is dropped and only one is chosen.

The percentage of rows with null or nan was 6.02% hence these were dropped using dropna function and the index was reset to include correct index count for every row. The remaining shape of the

dataframe consisted of 6 columns and 182954 rows. Using value_counts functions we can see the volume of different categorical elements and rename any categorical element with similar names.

The dataset is categorical, and we will use One Hot Encoding to turn categorical variables into numerical using either 0 or 1 for False and True. For better understanding view code script on notebook.

3. Methodology

In this section, I will include visualisations graphs for two attributes (JUNCTIONTYPE and ADDRTYPE).

3.1 Exploratory Data Analysis

We explore the data after it has been cleaned and ready for visualisation. We obtain two figure of bar graphs that illustrates severity level in junctions and address types.

The severity in junctions provides insight into where severity might be higher in certain junctions and where severity level doesn't incur much difference.

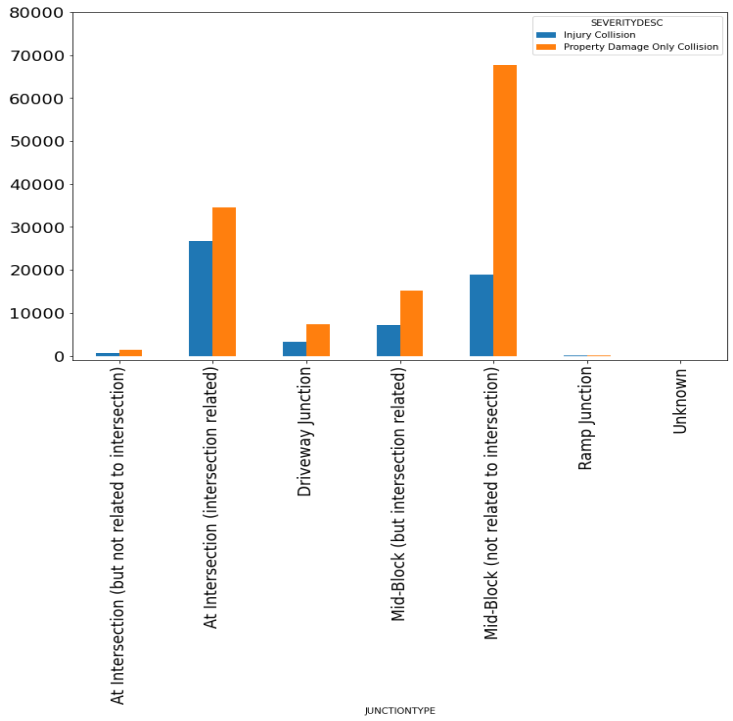

Figure 1. Number of collisions recorded in junctions with severity level

The address graph provides insight on the level of collisions occurring in each junction address type and the correlation between the severity level.

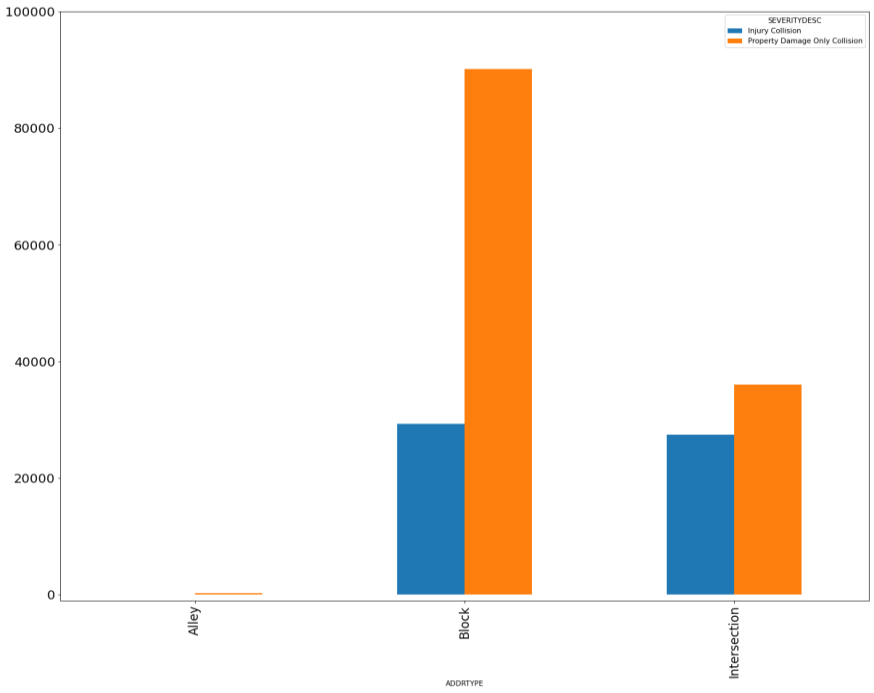

Figure 2. Severity level with respect to address type.

3.2 Model Development and Evaluation

The attributes for the dataset have categorical variables for their elements and the suitable model type for such dataset is a classification model. We will use Decision Tree model for our model. Decision Tree is a supervised learning technique that predicted the response by learning decision rules derived from features. It shows the possible outcomes and can be used to map out an algorithm that predicts the best choice mathematically.

We will use Jaccard Index and F1-score to evaluate our model. Jaccard Index, it is used to compare set of predicted values to their true values. The higher the score the better the accuracy. From 0 to 1, 1 being the best score. F1-score, it is the weighted average of the precision and recall. The score is between 0 and 1, best score is 1 and worst score at 0.

4. Results

The following results shown is an image of the decision tree with depth of 5 nodes and the evaluation results.

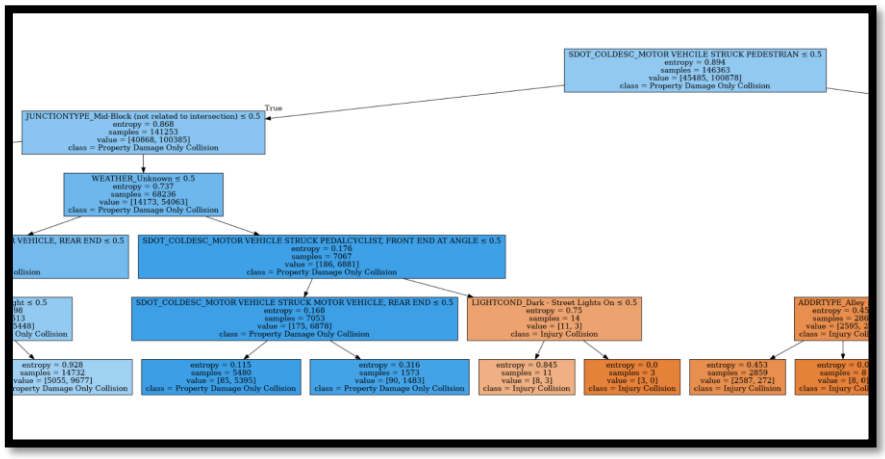

Figure 3. Decision Tree figure

Table 1. Evaluation

	Jaccard	F1-score
Decision Tree	0.73805	0.668051

5. Discussion and Conclusion

This project aimed to study the relationship between severity level of road traffic collisions in junction types. From our data analysis the graph shown in figures 1 and 2 illustrates that there are different levels of collision count in junction and address types. Such as the difference in severity level in junction types in not so much in all junction expect Mid-block (not related to intersection) junction type. As for address types, at block type we can see the same thing for Property Damage Collison Only but for Injury Collison between Alley and Block the difference is so much different.

To conclude, this project aimed at exploring the data to provide insight in severity levels for road collisions at junctions. The predictive model would be useful to help local authorities decide on whether to implement new safety measures in certain areas.