Car Accident Severity

Devansh Goel

September 7, 2020

Introduction

Road safety is considered one of the most important problem of the world. 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. And for this project, the aim is to build a model that can predict the severity of accidents in junctions around different types of address block and collision types using attributes such as weather and light conditions.

Target 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, and hence this could then lead to applying new road safety measures.

Data

The features used will be: • SEVERITYCODE • ADDRTYPE • JUNCTIONTYPE • SDOT_COLDESC (Description of the collision) • WEATHER • LIGHTCOND

The dataset has information gathered on the road traffic accidents of Seattle City. The initial dataset consists of 38 columns (features/attributes) and 194673 rows. Python packages will be used to conduct this study. The dataset will be cleaned according to the requirements of this project. Missing data information will either be substituted using valid means or dropped – considering the amount of missing data and the description of individual elements.

Methodology

For implementing the solution, I have used Github as a repository and running Jupyter Notebook to preprocess data and build Machine Learning models. Regarding coding, I have used Python and its popular packages such as Pandas, NumPy and Sklearn.
Once I have load data into Pandas Dataframe, used 'dtypes' attribute to check the feature names and their data types. Then I have selected the most important features to predict the severity of accidents in Seattle. Among all the features, the following features have the most influence in the accuracy of the predictions: ADDRTYPE • JUNCTIONTYPE • SDOT_COLDESC (Description of the collision) • WEATHER • LIGHTCOND

Also, as I mentioned earlier, "SEVERITYCODE" is the target variable.
I have run a value count on road ('ROADCOND') and weather condition ('WEATHER') to get ideas of the different road and weather conditions. I also have run a value count on light condition ('LIGHTCOND'), to see the breakdowns of accidents occurring during the different light conditions.
After importing necessary packages and splitting preprocessed data into test and train sets, for each machine learning model, I have built and evaluated the model

PLOTS
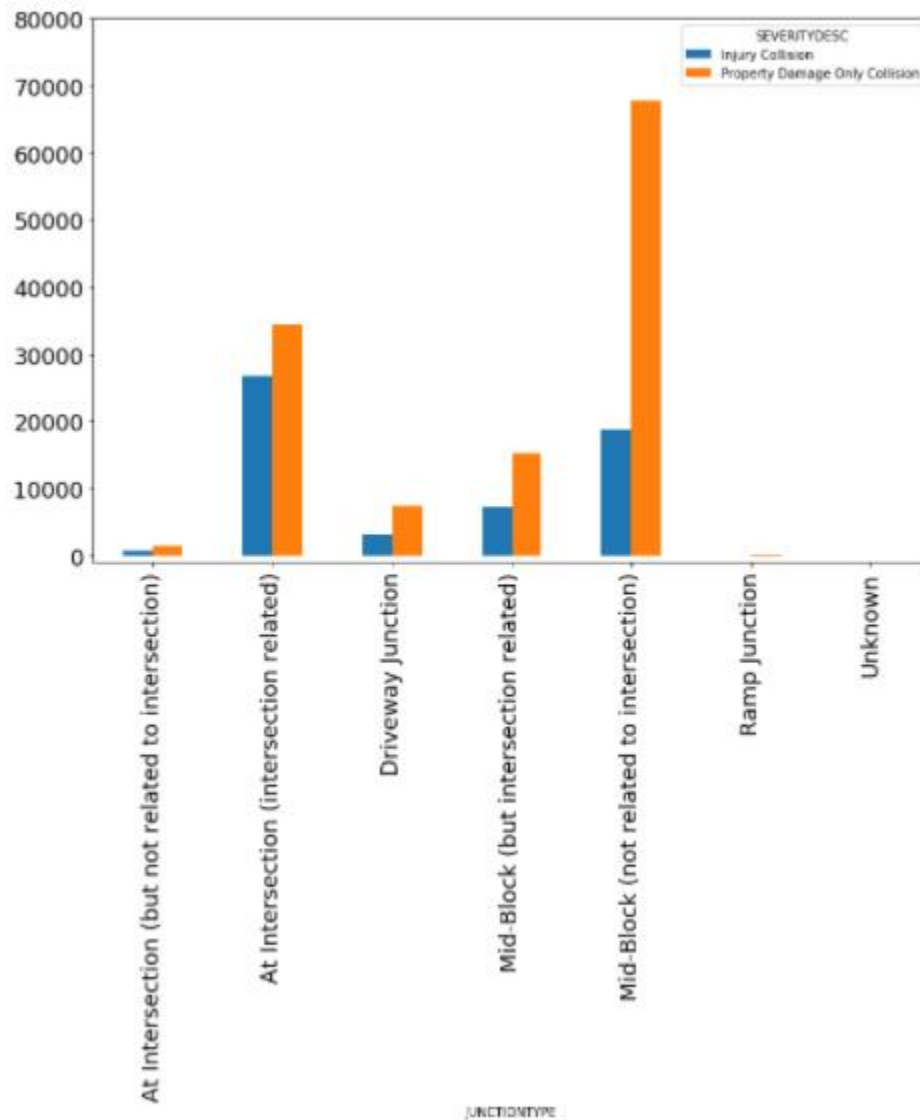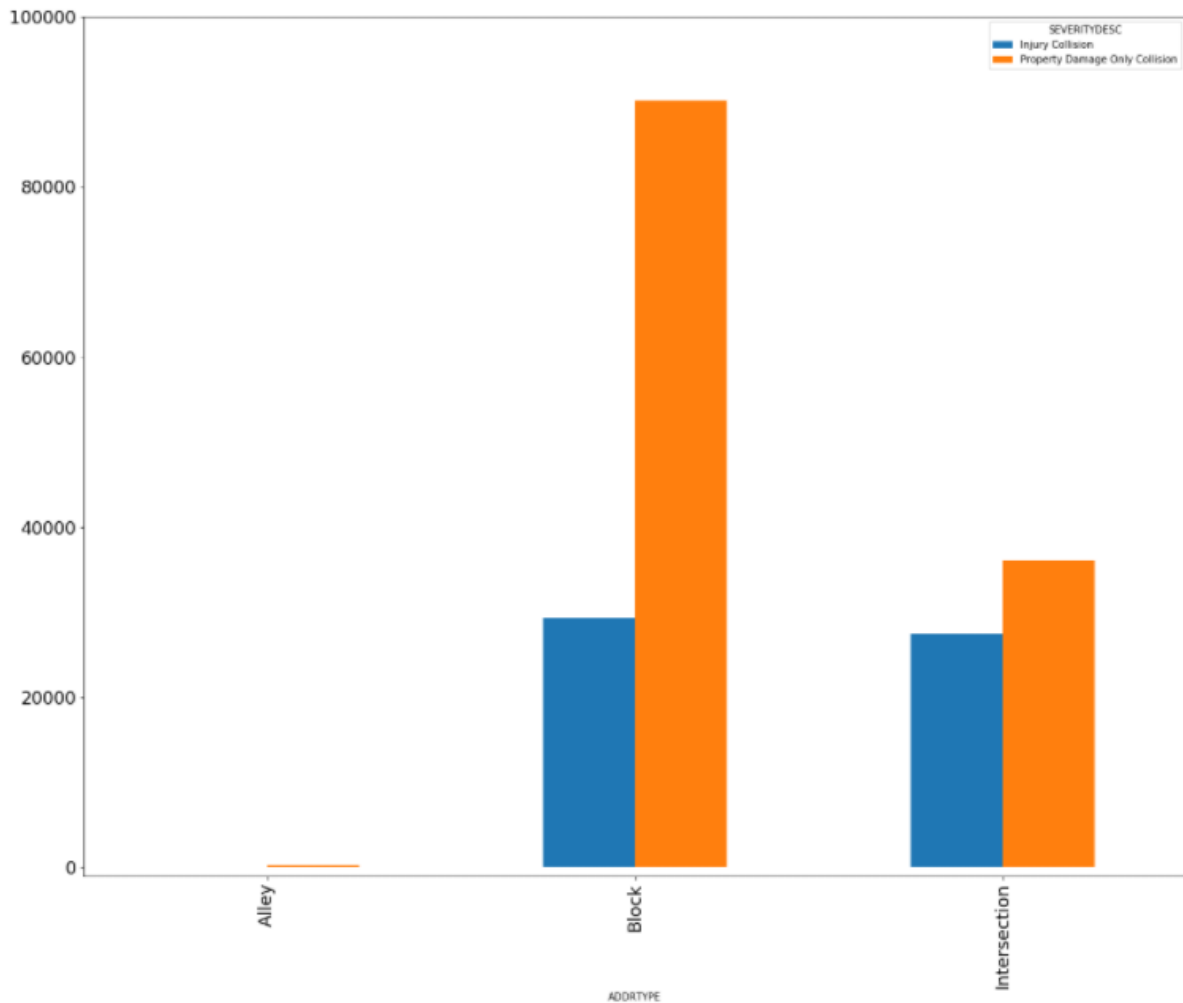
RESULTS:

From the decision tree analysis, it is shown that the accuracy of the model is `0.738`

Conclusion

Based on the dataset provided for this capstone from weather, road, and light conditions pointing to certain classes, we can conclude that particular conditions have a somewhat impact on whether or not travel could result in property damage (class 1) or injury (class 2).