September 12, 2020

Predicting the Severity of Car Accidents in Seattle

Introduction

There is no doubt that the reduction of tragic road crashes is a wish of many people. According to statistics posted on the WHO website, traffic collisions kill 1.35 million humans worldwide each year, with losses equivalent to 3% of GDP in many countries (https://www.who.int/health-topics/road-safety, accessed on 9/5/2020). With traffic accidents being the leading cause of death for children and young adults between the ages of 5 and 29, according to WHO statistics, there is an urgent need to address this issue.

The City of Seattle is also focusing on this issue, with the "Vision Zero" plan to reduce traffic fatalities to zero by 2030 being launched in 2015 (http://www.seattle.gov/visionzero, accessed on 9/5/2020). According to the "Vision Zero 2019 UPDATE" report issued by the city, traffic fatalities have shown a brief plunge from over 40 recorded in 2006, but have remained generally between 15 and 25 for more than a decade since then until the end of 2019, with a marked downward trend is not indicative (http://www.seattle.gov/Documents/Departments/SDOT/VisionZero/2019_VZ_Update_Report.pdf, accessed on 9/5/2020). This study analyzes traffic accident data from the City of Seattle and explores possible interventions and potential countermeasures to realize the Vision Zero plan, based on an examination of the conditions under which serious traffic accidents, such as personal injury, are occurred.

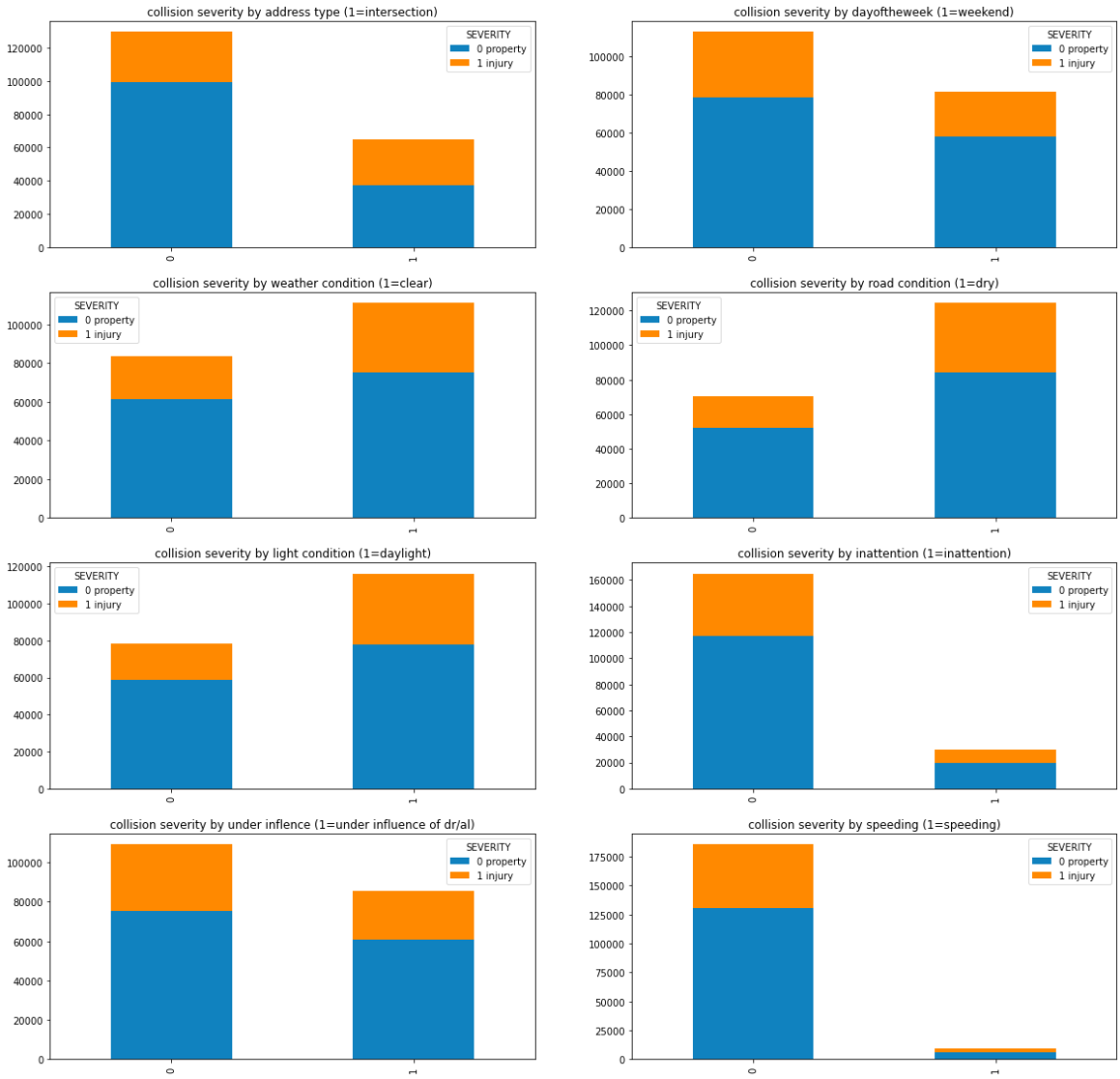
Data

The data used in this analysis are traffic accident records collected by the Seattle Police Department for the years 2006-2020 (N= 194,673). Cases labeled as property damage (N= 136,485) and personal injury (N= 58,188) were extracted. Here, we aim to gain insights into the mechanisms of serious road accidents by analyzing the characteristics linked to personal injury. The eight key features used for prediction are the address type of

the accident location (intersection or not), day of the week (weekend or not), weather (clear or not), road conditions (dry or not), daylight conditions (daylight or not), collision caused by inattention (yes or no), driver under influence of substances (yes or not), and collision caused by speeding(yes or no). These variables were examined for their recorded content, and binary-coded described as above (resulting in missing values being coded as 0). Features were finally standardized to mean 0 and standard deviation 1.

Methodology

The relationships between the dependent variable, i.e. accident severity, and these features are graphed as follows.



As can be read from the graphs, accidents are more likely to occur at non-intersections, during the weekdays, on sunny days, on dry roads, under the daylight and so on. It is likely that this result is due to the high volume of traffic and the number of pedestrians going out in such conditions, but injury ratios vary from situation to situation, and it is unclear what combinations of these features cause them. In this analysis, we apply machine learning models such as decision trees to examine how these features explain the outcome variables.

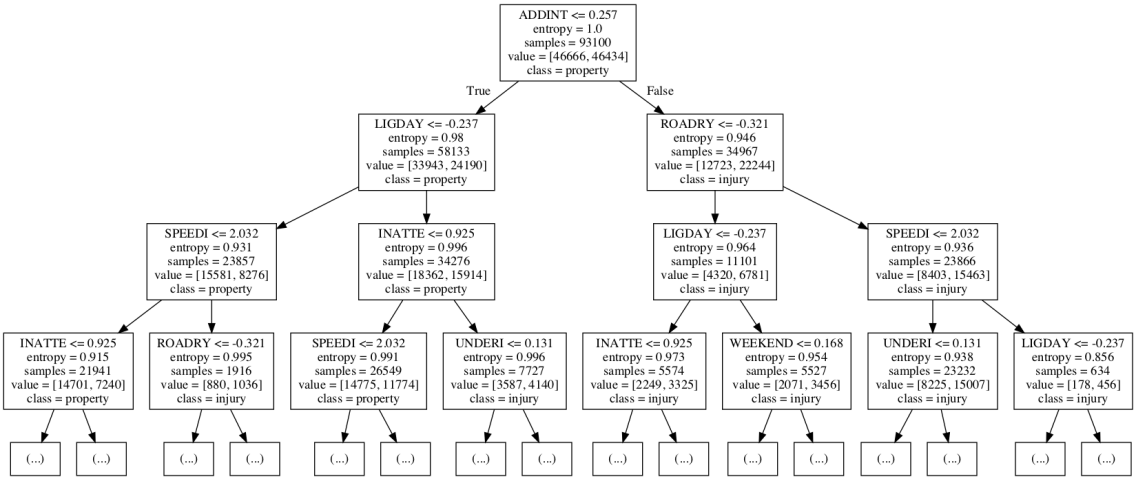
Since the number of property damage cases was much higher than the number of personal injury cases (Ns=136,485 vs. 58,188), and bias was expected to occur in the model estimation, a prior down-sampling was performed to match the number of property damage cases with the number of personal injury cases.　As a result, the total number of cases was 116,376.

Result

From a total of 116,376 data sets (58,188 each for property damage and personal injury), 20% were extracted as a test dataset and the rest were used as training datasets for machine learning. Support vector machine (SVM), logistic regression(LR) and decision tree(DT, depth=5) model were trained by accident severity as a target along with eight feature variables. The evaluation metrics for each of the three models for the test data are shown below.

	accuracy	F1 score	Jaccard score
SVM	0.6121	0.6121	0.4410
LR	0.6106	0.6102	0.4392
DT	0.6128	0.6128	0.4418

Although the prediction performance of all the models is not that high, the decision tree model has the highest predictive power in relative terms. Moreover, because the nature of the decision tree model provides the best clues for specific interventions, we adopt this model. The visualization of the decision tree up to a depth of 3 is shown below.



Discussion

The following points can be noted from the decision tree model:

- The number of personal injury accidents at intersections is very high.
- At intersections, personal injury accidents are particularly common when the road surface conditions are good and the speed is excessive.
- At non-intersections, personal injury accidents are rather common under sunlight and when the driver is paying attention.
- Also at non-intersections, personal injury accidents due to excessive speed are more common under non-sunlight conditions.

From these results, first of all, countermeasures should be taken at intersections, where traffic accidents occur frequently. In particular, accidents caused by excessive speeding are common when the road environment is good, and countermeasures such as traffic signal control, speed restrictions, and speed enforcement should be considered in this situation.

On the other hand, at non-intersections, many accidents occur under sunlight and when drivers are paying attention, and it may be more effective to alert pedestrians. Also, at non-intersections, many accidents occur due to excessive speed when not in sunlight, and speed restrictions and speed enforcement may be effective in this regard as well.

Conclusion

While the above recommendations can be made from this analysis, the first problem is that the model is not very accurate. One possibility is the lack of important features that were available, and more in-depth consideration of the mechanisms of road accidents would be needed to compensate for the missing indicators. In addition, regional characteristics are mainly dealt with only the shape of the accident site (e.g., intersection, alley/block), and more consideration should be paid to the geographical conditions specific to the accident site. The reduction of road accidents is an important issue that many people aspire to, and further study will be required.