Predicting Accident Severity

Rosa Ferrusca

October 2020

Introduction

Background

According to the World Health Organization, road injuries are within the 10 top causes of death worldwide, occupying 8th place in 2016. “Approximately 1.35 million people die each year as a result of road traffic crashes” (WHO, 2020). Also, numbers indicate that around 20 to 50 million suffer injuries and disabilities, turning this problem not only a death cause but economic and social problem around the world. As a result of car accidents many people go through time spending, pain, family member losses and economic losses.

Although there are several causes of car accidents, what is certain is that this is a huge problem that affects all countries around the world, to which governments have paid attention and organizations worldwide support. Efforts go from increasing traffic laws and requirements to improving road infrastructures and people consciousness, however there are also unmeasured consequences for people driving and passing through existing car accidents that result in huge waiting times and even in new accidents nearby that could be avoid.

To address this problem, we will look into car collision information and search for possibilities to predict car accident severity so this data can be used by governments or organizations in local communities, to provide live information on actual car accidents for people to take precautions or change their mobility decisions, avoiding time spending or possible new accidents cause by existing ones.

As part of government efforts to reduce deaths and economic impacts due to car accidents, this should be very helpful for local administrations to help reduce secondary impacts and new accidents and delays for their citizens.

Data

We will be using data from Seattle city with collision data from 2004 to 2020, looking into different variables such as light, road and climate conditions which are changing circumstances that affect the severity of accidents, thus changing the affectation to the nearby roads and traffic.

Data set was recorded by Traffic Records Group and provided by Seattle Police Department. It contains 194,673 records with 38 attributes.

Methodology

For selecting the data that could be more useful in the analysis, data was analyzed and cleaned up. First, one column was duplicated so it was deleted. After that, data such as specific addresses, incident and report number, and codes and descriptions which are non-related to severity were also deleted. Final data set included attributes that affect drivers and people around car accidents, such as light and weather conditions and attributes that

are characteristics of the accident that might have relationship to the severity, such as car and people involved and collision type. Total attributes kept for further analysis were 13 out of the initial 38.

As much of the attributes are descriptions, data was modified to numerical values to be able to look into relationships among them. I started by analyzing ratios to see what attributes had more relationship to severity.

For each attribute, frequency was plot to drop all the rows with no useful information. After analyzing the 13 attributes and how much representative data they provided for the accident severity, data was reduced to 143,561 records with 7 attributes with highest impact in information related to severity.

With that final dataset and all data in useful data type, training and testing data was set to build a model that can help predict accident severity. As there is historical information for different independent variables and our target variable has only two options, logistic regression was the model selected to predict a binary variable, in this case severity code 1 “property damage” or severity code 2 “injury”.

After running the model, a 68% accuracy was obtained and data showed that most of the accidents occur in daylight, with dry roads, clear weather and during the morning.

Results

After all data cleanup and analysis, findings showed that car accidents’ conditions are mostly the same for a severity that causes property damage and the ones ending in injury. Nonetheless, injuries might cause bigger consequences for surrounding areas and post-accident traffic. Therefore, the interest in finding a model with accepted accuracy to predict which severity the accident will have based on the different independent conditions so that citizens driving or passing near the accident can take timely decisions to change their routes.

Discussion

Even though all the final variables selected were all similar for the two severity types, there were 2 conditions that showed higher incidents that might caused injuries and be a good indicator for citizens to avoid the area. These were accidents in intersections and during afternoon.

Unfortunately, for this analysis there wasn’t enough data on speed, which we thought could be a good variable to help better differentiate between an accident causing severe injuries.

Conclusion

Based on the results showed by this study, a prediction model with a 68% accuracy is good enough for our purpose of helping local governments inform their citizens with live information on actual car accidents, for them to take precautions and make better mobility decisions every day. This way, local communities have better results on efforts of avoiding increasing car accidents.

References

WHO. (2020). Road traffic injuries. October, 2020, de World Health Organization Sitio web: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries