Severity of Car Accidents

Abdulla Eid

August 30, 2020

1. Introduction

1.1 Background

Car accidents cost millions of lives loss and billions of money loss every year. The severity of a car accident is the most important factor in categorizing car accidents. The loss in lives and money depends on the severity of the car accident. Reducing the severity of car accidents will save a lot year by year and will make the car automobile more safe. It is of great interest to find ways to reduce the severity of car accidents by taking precautions that depend on many factors like road and weather conditions which could help reducing the severity of an accident.

1.2 Problem

Data that might contribute to determining the severity of a car accident will include the location of the accident, weather conditions, car speeding, road conditions, number of people and vehicles involved in the accident. This project aims to predict the severity of a car accident given these many variables and attribute.

1.3 Interest

Obviously, traffic control authorities would be very interested in accurate prediction of the severity of accidents in order to take measures that reduce the severities of accidents, e.g., diverting some roads in bad weather conditions or reducing the speed limits in some highways. Others such as drivers, medical teams may also be interested from their perspectives.

2. Data acquisition and cleaning

2.1 Data Sources

The data for this project has been provided by the Seattle Police Department and recorded by Traffic Records which has been shared to us via link by Coursera. A detailed data base of car accidents that happened in Seattle from 2004 till now can be found here. There are 194673 observations in the file with a target attribute is the severity of the accident which is categorial variable with only two values, 1 for moderate and 2 for severe.

2.2 Data Cleaning

A quick overview from the data downloaded from the link above, we found some features have a lot of missing values and so these are dropped, e.g., speeding is only recorded on 9333 incidents. Also observations that are lack in the major features such as RECORDID are dropped. This left us with a shorter list but rich in information to be drawn from.

2.3 Feature Selection

After data cleaning, there were 38 features in the data. Upon examining the meaning of each feature, it was clear that there were some features are irrelevant and does not contribute to the target variable. For example, there was feature of the incident ID or report no. Also there is data related to pedestrians involved in the accidents and looking at these data, there is little information that could contribute to the target label. After examine the data for correlation, we decided to keep the following important features

Feature	Meaning
SEVERITYCODE	The severity of the accident
X	The x coordinate of the accident
Y	The y coordinate of the accident
ROADCOND	The condition of the road
WEATHER	The weather condition
JUNCTIONTYPE	The type of the junction where the accident took place
PERSONCOUNT	No of people involved
VEHCOUNT	
INCDATE	Date of the accident
ADDTYPE	The Type of the address (Alley, Block, Intersection)
COLLISIONTYPE	Collision Type

3. Exploratory Data Analysis

3.1 The target variable

The target variable is SEVERITYCODE which is a given attribute and can take two numeric values, one for moderate injuries and two for severe injuries. Despite these values, the attribute is a categorial variables with two binary values and will be treated as such during this project. In Figure 1, the distribution of the values is given.

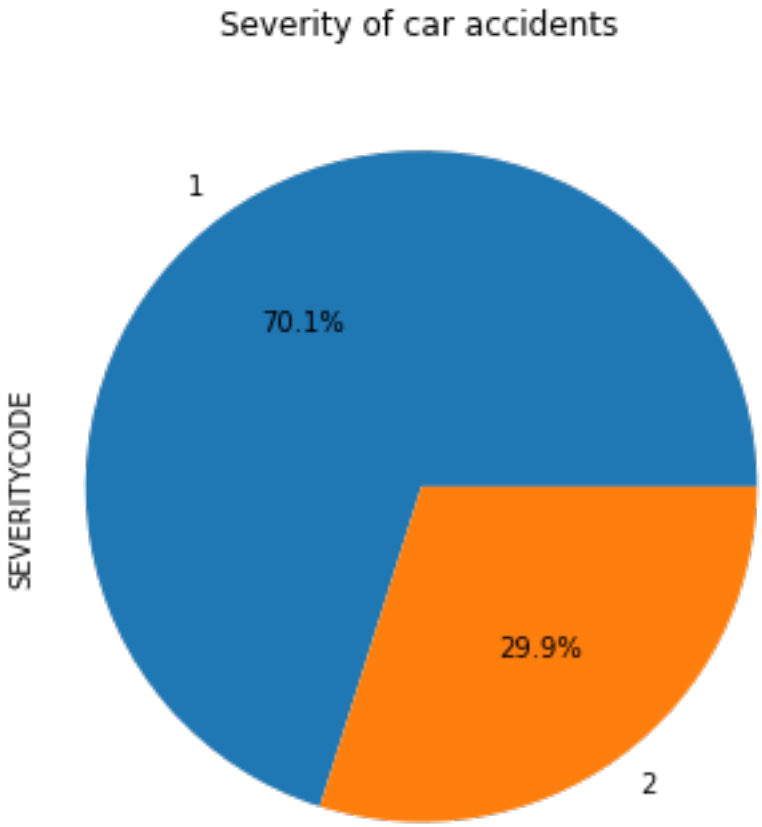

Figure 1 Moderate and Severe Injuries Distributions

3.2 Relationship between Severity and Road Conditions

As in Figure 2, we notice that the most two road conditions with incidents are Wet(30%) and Dry(70%). To some extend, we would expect more sever injuries associated with the wet road conditions.

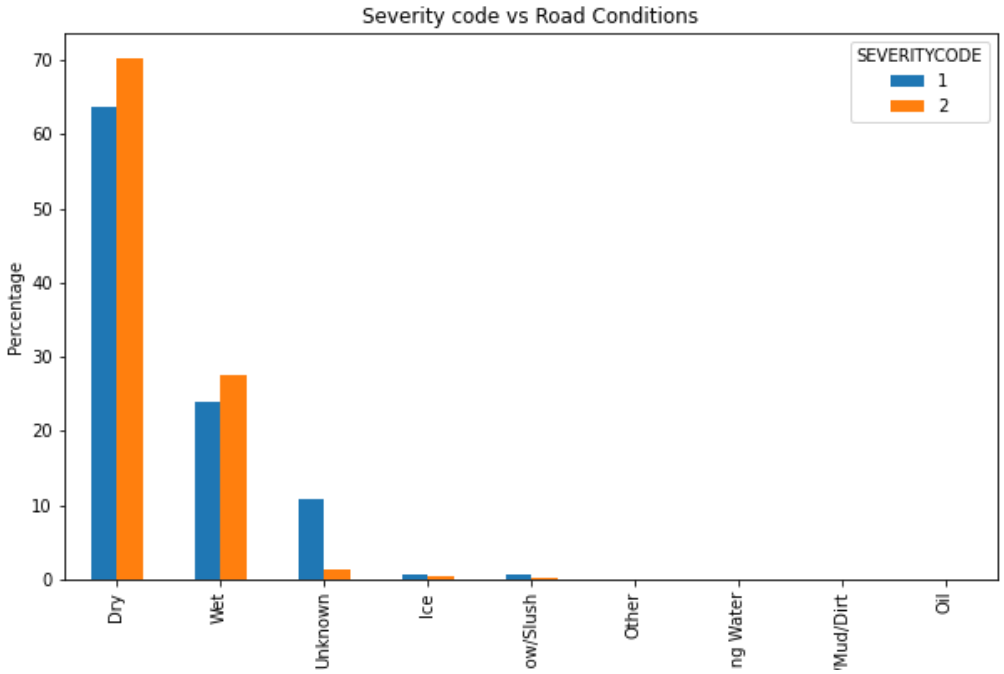

Figure 2 Percentages of road conditions vs severity

3.3 Relationship between Severity and Weather Conditions

As in Figure 3, the sever injuries happen in the clear weather (60%) while that in the raining or overcast is only accountable for 40%. The rest of conditions are not really contribute to severe injuries. Something to note here is that in extreme weather conditions (raining or overcast), the severe injury is more likely to happen than the moderate injury.

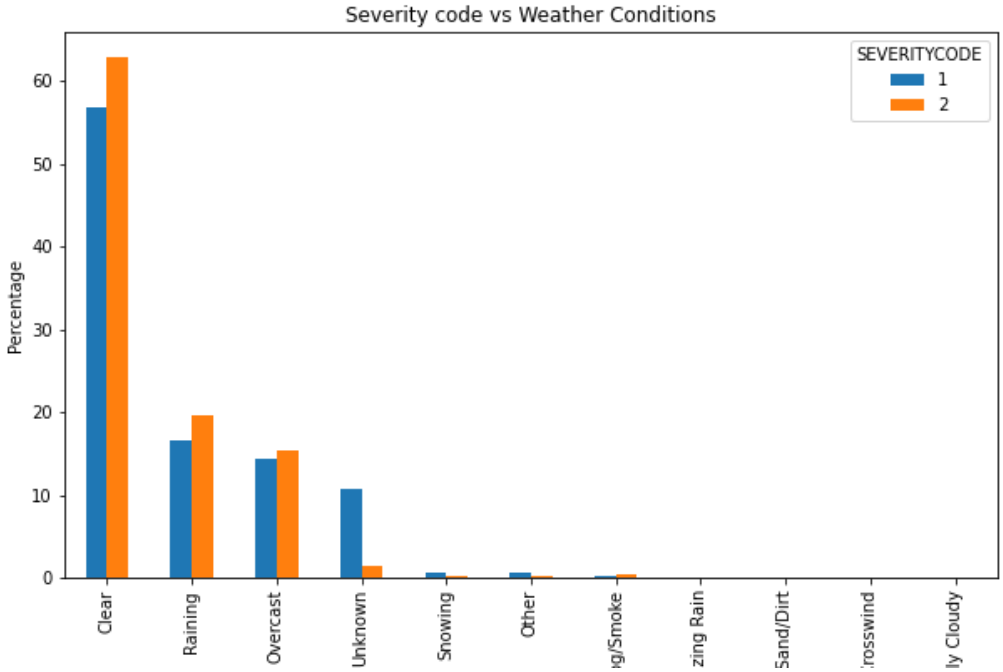

Figure 3 Percentages of weather conditions vs severity

3.4 Relationship between Severity and Junction Types

This is an interesting observation in Figure 4, where severe injuries occur in intersection and intersection related junction or in the driveway. This gives a hint that more focus and regulation must be given to intersections. As expected injuries of incidents in mid block are moderate.

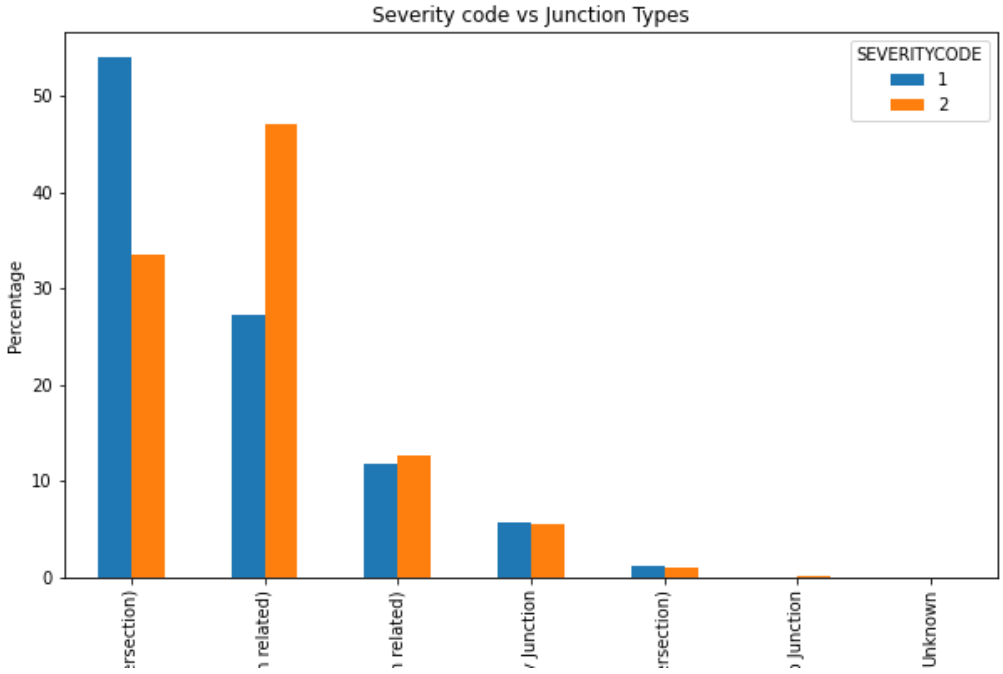

Figure 4 Percentages of junction types vs severity

3.5 Relationship between Severity and Collision Type

Accidents involve pedestrians and bicyclists result in severe injuries. In addition, rear-end and angles accidents are also sever accidents.

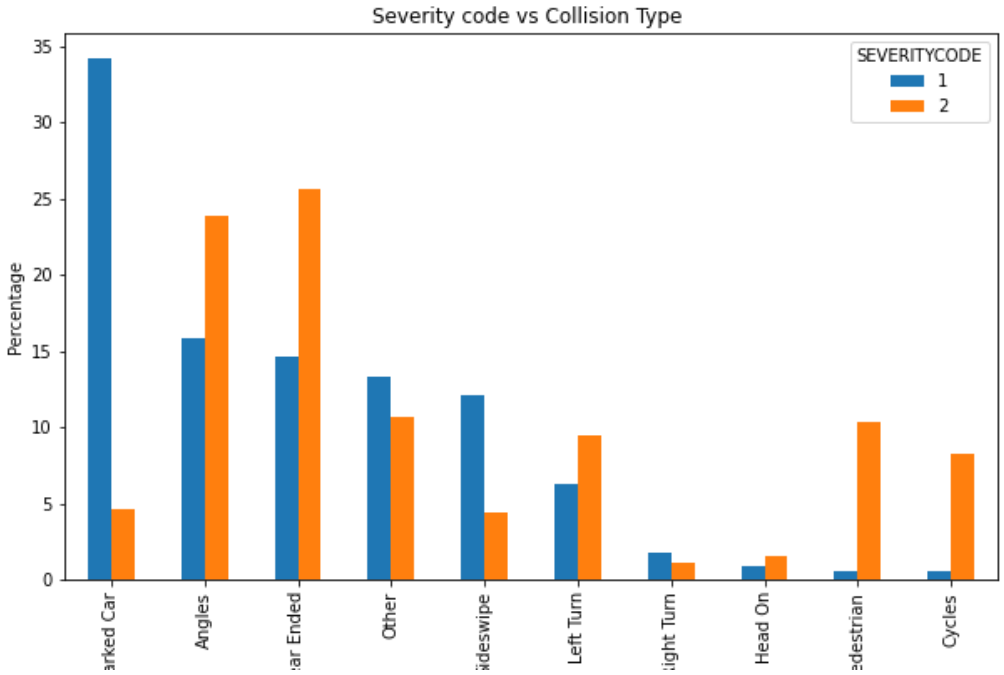

Figure 5 Percentages of collusion types vs severity

4. Predicted Models

The predicted model that was used here are the classification models since our target variable is categorial variable. The dataset was split into 80% to be used as training set and 20% as test set. 3 main models were built and trained using the data set with the attribute mentioned in Section 3, namely K-nearest neighbor, decision tree, and logistic regression. For each model, Jaccard, F1-score metric were calculated and presented in Figure 6. The KNN model was the best as it gives high accuracy rate (with K=6 and accuracy rate=0.744).

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.744266	0.69749	NA
Decision Tree	0.752459	0.689341	NA
SVM	NA	NA	NA
LogisticRegression	0.702941	0.585244	0.589575

Figure 6 The Accuracy of the train model using different ML Algorithms

5. Results and discussion

Although, there are many factors contribute to the severity of injuries, few features can be highlighted. For example, it seems that intersections are the most places where injuries are severe. Also, the wet road condition, rainy or overcast weather contribute more into severity of accidents. Also, the collision location of the car is attributed to the severity of the accident, e.g., most rear accidents ended up in severe injuries. This could help car manufacture to focus on that part of the car and increase the strength of it.

6. Conclusion

The purpose of this analyze the relationship between the severity of an accident based on the several variables. The weather condition, road conditions, junction type, and collision type among the most important features that affect the severity of an accident. I built classification models to predict the severity of an accident. These models can be very useful in helping the traffic directorate to make good recommendations and actions to prevent severity of injuries in car accidents.

7. Future Directions

In this analysis, only four attributes were on great interest and have direct relation with the target variable. Other attributes could possible affect the target variable but unfortunately, there was no enough data to explore these in the data sets. It will be of highly interest to record the speed of the vehicles just before the accidents which could help reengineer the roads by adding extra ramp, reducing the speed limit, or make amendments to the roads infrastructure.