**Predicting the Severity of Car Accidents**

By:

Brian Haught

**1 Introduction**

**1.1 Background**

Transportation is a major part of the economy and our everyday lives. From shipping to baseball games to driving to work, everybody uses transportation. One of the major pitfalls of transportation is accidents and their severity. Due to the nature of vehicle accidents, they can be very perilous. Because this is such a major part of our lives, we should make every effort to prevent accidents through any means necessary to lessen or eliminate these accidents and their severity.

**1.2 Problem**

In this project, we will provide information to the traffic engineers to identify areas of local hotspots in Seattle to evaluate if they need speed adjustments and infrastructure based on conditions. These speed and infrastructure adjustments will be based upon weather, light, and road conditions. We will use clustering to determine these hotspots and identify an algorithm that will determine severity of accidents based on these conditions so they will be able to adjust the speed limit accordingly.

We will also include driver based contributions such as inattention and underinflated tires to ensure that conditions are not misinterpreted.

To identify these locations, we will use density-based clustering (DBSCAN) to identify hotspots in Seattle. DBSCAN groups points that are closely packed together and marks points outside of these groups as noise. Therefore, using this algorithm, locations in which a high density of accidents take place will be highlighted as clusters. We can then plot the location of these clusters using folium.

**2 Data**

The data we are using is based on police reports from Seattle, Washington. As such, lesser accidents that did not require police assistance are not included in the data. The feature we are going to measure from the data set is the severity of the accidents which is represented in "SeverityCode" which has a value of 1 or 2. The Features we are going to use for causation are Speeding, Weather conditions, Road conditions, Light conditions, Driver inattention, and Underinflated tires.

**2.1 Cleaning**

The data has been cleaned up to make it more usable. As such a few assumptions had to be made. For instance, in the “Speeding” information, if no information was provided it was assumed to be not speeding. This assumption was repeated for inattention and underinflated tires.

Another issue that arose was location was not always provided. As this study is heavily dependent on location, all reports that did not include a location were dropped and not considered.

**3 Methodology**

After cleaning the data, there were 189,339 samples and 38 features. Most features were redundant descriptions of other features. The data was split into two basic categories. Location data and causation data. Certain features that are not preventable regardless of infrastructure and speed were not included.

After considering the possible features, the final choices were Weather, Road Conditions, Light Conditions, Inattention by driver, and Underinflated tires.

**3.1 Predictive Modeling**

By using machine learning, we can create models that can predict the severity of accidents based on the provided features. I created and evaluated several models using two types of machine learning. The two types looked at were regression and classification. Regression models included linear regression and support vector machines. Classification models included K-Nearest Neighbors and decision tree. These models were run using an 80/20 split of the existing data.

The models for both classification and regression had similar accuracies. By running both the Jaccard score and the F1 score, we can get a good idea of the accuracy.

Table

Description automatically generated

**4 Results**

To highlight the causation of each individual Feature, we plot the severity of accidents for each **Chart

Description automatically generatedChart

Description automatically generated**

**Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated**

Using the barcharts above, we can see that the individual Features did not have a significant impact on accidents. This shows that it is a combination of Features that contribute to accidents. Traffic engineers will use these combinations to design roads and equipment to assist in the prevention of accidents.

The largest number of accidents were severity level 1 and involved parked cars.

**5 Discussion**

The main conclusions from the given data are that the majority of accidents were not due to specific conditions of light, road conditions, light conditions, or speeding. The majority of accidents were due to hitting parked cars. As such, the most focus by the engineers should be on improving or adding parking structures or evaluating why the parked cars were hit. By fixing this error, greater than 40% of the accidents could be eliminated all together.

**6 Conclusion**

In conclusion, there are many ways to improve infrastructure to prevent accidents or reduce the severity of accidents. The majority of which can be improved through the evaluation of parking and/or adding parking structures to eliminate the majority of accidents that involved parked cars. Other evaluation points could be minimizing accidents by dropping speed limits. Unfortunately, accidents are an expected part of driving and dropping the speed limit to the point where there are no accidents would mean that we would get nowhere. There must be an acceptable level of accidents or an acceptable level of severity of accdients.