PREDICTING CAR ACCIDENTS SEVERITY

David Bertel

Oct 2020

1. Introduction
   1. Background

The city of Seattle has been gathering information about accidents and their severities. It has by now a dataset with these accidents and some features associated to the accidents. These features may be used to identify correlation or causation depending on the features in the dataset, which we are going to investigate in this report.

* 1. Problem

The worst outcome of an accident is casualties. Even if there are no casualties, normally when an accident occurs the people involved in it have to wait for the police to clarify the facts before being able to remove the cars. This process causes delays in the traffic.

* 1. Interest

All people using a car are subject of traffic delays and therefore would benefit if there is an improve in the traffic i.e. if it would be possible to predict delays and best routes depending on certain features. On the other hand, companies developing autonomous cars can benefit from the outcomes of this study since it would help the to identify high risk situation or areas. This information about risk would be very valuable for insurance companies.

Last but not least, being able to predict the severity of accidents on certain areas may be useful for the city of Seattle to deploy emergency services or even a proper traffic infrastructure.

1. Data acquisition and cleaning

Before starting with any kind of predictions it is important to know the dataset and the format of the information contained in it. Although it is not recommended by some data scientists, because it could bring biases to the analysis, a preliminary visualization of the data helps to have a first idea of the problem. Some important features that are going to be used are described in text. Using proper transformation, I converted them into *int*

*Table

Description automatically generated*

Table

Description automatically generated

* 1. Features

After analyzing the information in the dataset, I concluded that reducing the prediction of the severity of an accident can be reduced to some particular conditions: WEATHER, ROADCOND and LIGTHTCOND. These features represent the weather, road condition and light condition respectively.

* 1. Normalizing the data

To avoid some features having more importance than other in postprocessing I used the *sklearn* to normalize the data.

1. Data analysis and results.

To analyze the data set I used the K- Nearest neighbor method. By reducing the prediction to 3 features I reached an accuracy of the model of around 70%. This information would surely help all the stakeholders interested in the prediction.

1. Data visualization

Having the information about accidents and their severity, it is important to know where these accidents actually take place. The location is used to concentrate the measures to take specifically in these places. Most of the accidents occur in the city center. One approach to use the data would be that given certain conditions of road, light and weather the driver should transit with precaution in these areas or try to avoid them.

In the photo I limited the number of incidents to 500. It should be statistically representative of the distribution of accidents in the different areas of the city.

Map

Description automatically generated