# Car accident severity on Seattle.

Final report.

Michael Rodriguez Gamboa.

September, 2020.

## Introduction.

Can we predict the outcome of a transit accident before it happen? How severe would be? It's a job for a prophet? Well, no. With enough data, data analysis and Machine learning it's possible to know the result before the accident. If you can predict the severity you will help to the transit police, the pertinent authorities and the drivers to prevent accidents when dangerous conditions arise.

For this work I used a dataset from the weekly transit collision reports of the city of Seattle, USA. These Data contain all the relevant information on transit accidents from 2004 to 2020. I used it to build a predictive model with machine learning.

## Data.

The data used in the analysis is a CVS file with 194,674 reported accidents from 2004 to 2020 in the city of Seattle, USA. The file contains different important details from the incident like: localization, number of cars involved, number of people, condition of the road, light, weather, accident severity and others. The details vary from numerical values, descriptions and categorical classifications. All the columns in the file sum 39. It means that the dimensions in the initial dataset are: 194,674/39.

At first hand, the dataset had missing values in different column, and for that reason I start the analysis dropping the columns with more that 65% of missing data. Others columns with less empty values where filled with the most frequent value, to use it in the data frame.

Then I convert the categorical variables into dummies values,

The third step in the work was the statistical descriptive analysis, to find the relations between the variables. At this point I decided to use the "severity code" as the target or dependant variable, and the features or independent variables as the predictors for the model.

The most important features related to the severity code were:

|  |  |  |  |
| --- | --- | --- | --- |
| Features | Person Coefficient & P-Value | | Description |
| Persons Count | Coef= 0.130 | p= 0.0 | Total of people in the accident. |
| Pedestrian Count | Coef= 0.246 | p= 0.0 | Total of pedestrians in the accident. |
| Bicycles Count | Coef= 0.214 | p= 0.0 | Total of cyclists. |
| Vehicles Count | Coef= -0.054 | p= 8.177e-129 | Total of cars. |
| Location-Alley | Coef= -0.025 | p= 5.210e-30 | Where the accident occurs. |
| Location-Block | Coef= -0.195 | p= 0.0 | Where the accident occurs. |
| Location-Intersection | Coef= 0.199 | p= 0.0 | Where the accident occurs. |
| Junction-Intersection | Coef= -0.200 | p= 0.0 | Type of junctions near the accident. |
| Junction- Mid Block | Coef= 0.200 | p= 0.0 | Type of junctions near the accident. |
| Bad Weather- High | Coef= 0.038 | p= 1.779e-63 | Bad weather during the accident. |
| Bad Weather- Medium | Coef= 0.014 | p= 7.730e-11 | Bad weather during the accident. |
| Bad Weather-Very high | Coef= -0.016 | p= 3.828e-13 | Bad weather during the accident. |
| Road condition-Extreme | Coef= -0.011 | p= 2.907e-07 | Bad Road condition. |
| Road Condition- High | Coef= -0.020 | p= 2.176e-19 | Bad Road condition. |
| Road Condition-Medium | Coef= 0.040 | p= 7.743e-73 | Bad Road condition. |
| Darkness- High | Coef= -0.015 | p= 1.692e-12 | Darkness level. |
| Darkness- Low | Coef= 0.014 | p= 4.449e-10 | Darkness level. |
| At intersection? | Coef= 0.200 | p= 0.0 | The accident where at intersection? |
| Under drugs influence? | Coef= 0.044 | p= 1.90e-85 | The driver was on drugs? |

\*19 features selected as predictors.

## Methodology.

The methodology used

## Results.

In the model

## Discussion.

This can

## Conclusions.

Should be