**Car accident severity prediction model**

Akshit Jain

September 15, 2020

1. Introduction

We live in a world where a lot of statistic that directly affect road transportation are publicly available, these statistic include weather reports, visibility, congestion on certain days or dates, and so on.

These statistics are used to design a model which uses the available statistics in order to predict the likelihood of an accident happening.

Since almost everybody finds themselves on the road almost every day, this model is targeted to everyone. So when every a person wants to go out, they can check and see if it is safe for them.

1. Data

The dataset used contains data about car accidents provided by SPD and recorded by Traffic Records. I contains 40 columns, 39 of which are independent variable and a target variable 'SEVERITYCODE' which will be predicted. After the data was imported to the environment for training, it was discovered that the dataset set was imbalanced with a lot of unknown and null values.

The data was cleaned by assigning the modal value of each variable to the unknown and null values. Next, the variables were encoded using a label encoder to assign integer values to them.

1. Methodology

The balanced dataset is used to train three models which are: Logistic Regression, K Nearest neighbor and Support Vector machine.

* Logistic Regression: the model was trained and it attained an accuracy of 0.408.
* K Nearest neighbor: the model was trained and it attained an accuracy of 0.412.
* Support Vector machine: the model was trained and it attained an accuracy of 0.444.

1. Discussion

After training three different models, the k nearest neighbor had an accuracy of 41.2% and an F1 score of 0.396, the Logistic regression model had an accuracy of 40.8% and an F1 score of 0.356 while the SVM had an accuracy of 44.4% and an F1 score of 0.393.

1. Conclusion
2. The models trained above are a way of predicting whether or not the will be a collision on a certain day given the conditions. For cases where there will be collision, the model goes further to predict how severe the collision would be. This model is great and useful to a lot of people.