

Car Accident severity Project

Applied Data Science Capstone

Introduction

The surprising weather conditions in Seattle is one of the most important motivations for doing this project. The weather can change drastically from sunny to rainy or snowy!

For this project, I employed the old accident record and its severity along with other information such as weather conditions, geographical conditions, number of pedestrians, number of injured people in car accident, time of the day, etc. I applied machine learning algorithm to predict the severity of the any future accidents based on mentioned features.

The aim of this study is to determine under which weather conditions (and other parameters) a severe accident is more likely to occur, so we can prevent the accident to be happened! In this case we can warn the drivers to avoid driving under certain conditions and use a safer road instead.

Data

It is a large dataset with a dimension of 193673 x 38 (37 attributes are available in this dataset). The severity data is imbalanced as 126485 rows belongs to group 1 severitycode and 58188 rows belongs to group 2. This severity column can be used to train and predict the model.

Methodology

In this section the methodology of this project is well explained:

- At first, the necessary features are selected for the aim of machine learning.
- Replace NAN values with the Mode of each feature.
- Convert categorical data to Dummy Variables.
- Split data to train and test
- Train decision tree classifier model to predict accident severity
- Determine the f1 score

Results and Discussion

In this study we employed a Decisiontree classifier method to predict the severity of the accidents based on some other features such as weather conditions, road condition, time or the day, and so on. Results showed that the f1 score for the train data is 0.89 which decreases to 0.81 for the test data. As a result, the model performed well to predict the severity of the accidents with high accuracy. Thus, the Proper warning can be sent to drivers ahead of time to prevent possible accidents.

Conclusion

The aim of this project was to get hands on experience on various phase of the data science projects such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment.