Prediction Modelling of Severity of Road Accident.

Applied Data Science Capstone Project offered by IBM.

Himanshu

September,2020

1. Introduction

1.1 Background

Road Accidents are one of the most prevalent issues of Modern Society which poses a heavy toll of losses to mankind. WHO describes the road traffic system as the most complex and dangerous system with which people have to deal every day. As per Global status report on traffic safety, 2018 there are 1.35 million traffic deaths in the world occurring annually. Also, the report suggests that Road traffic accidents rank 1st for the cause of deaths of people in the 5-19 years of age and ranks 8th for people of all ages. Hence, Accident prediction becomes the need of an hour for the optimization of public transportation, enabling safer routes, and cost-effectively improving transportation infrastructure; all leading to safer roads. Due to it's high significance, the topic has attracted attention of many researchers since the last decade. Analyzing the impact of various events such as traffic events(e.g., congestion, construction and road hazards), weather(e.g., temperature, visibility and wind speed), points-of-interest(e.g., traffic signal, stop sign and junction) using the availability of Big data in modern day with utilisation of various Advanced Statistical Methods have ease the thought process and lead the researchers to form a Predictive Model more efficiently.

1.2 Scope

The Goal of the project is to create a Predictive Model for Severity of Traffic Accidents. The target or response variable is "Severity". The Case Study involves the study of the City of Washington, Seattle. The datasets and all other required information are obtained from the Seattle GeoData website:
https://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0/data?selectedAttribute=ADDRTYPE
The project is based on Cross Industry Process for Data Mining(CRISP) Methodology. As it is a case specific project, CRISP will provide a clear intent for the same.

The following six steps will be followed throughout the project:

A. **Business Understanding**: Intention of Project is to create a Predictive Model to predict the “Severity” of the accident based on the dataset provided. Hence it is clear that a Supervised Machine Learning Model will be advisable to use. The clarity of “labeled data” will further help in creating a training and testing data.
B. **Data Understanding**: The Data as mentioned above is obtained from the Seattle website. This step will involve understanding the various aspects of Data obtained and a deep analysis for the same.
C. **Data Preparation**: Based on the results from the previous step, the raw data will be transformed into a usable subset. Once the dataset is chosen, it must be checked for questionable, missing or ambiguous cases.
D. **Modelling:** In this step, data will be expressed through whatever appropriate models provide meaningful insights and any new knowledge if present. Different patterns and structure of the data will be captured using the various Modelling techniques.
E. **Evaluation**: The Model selected will be tested. A pre-selected test set will be used on the trained model. The effectiveness of the model will be analysed. Based on the analysis results efficacy of the model will be determined.
F. **Deployment**: Deployment for the model presented will be done on Github as part of the Course Project Guidelines.