

Coursera Applied Data Science Capstone Project

Accident Severity Prediction



August 28, 2020

Author: IFTEKHAR ALAM

Capstone Project:

Introduction/Business Problem:

Traffic accidents are a significant source of deaths, injuries, property damage, and a major concern for public health and traffic safety. Accidents are also a major cause of traffic congestion and delay.

The safety impact of weather and road conditions on the traffic accident causality need to be analysed in order to build a model to predict the probability and severity of the accident. This would act as a warning to the driver so that they can drive more safely or change the travel plan if possible.

Imagine you getting into your car and head unit of the car is auto paired with to your mobile phone. The head unit collect the data based on the mobile tower location and then get the current weather and road condition. The information obtained from the weather and road condition will then be used to predict the severity of the accident and so that you drive carefully.

Fortunately, Advanced data analysis and machine learning will help taking this decision with the information available in abundance from the traffic department of the city. The Data collected from the Seattle Department of Traffic (SDOT) has records of all accidents that occurred during time period 2004 till 2020. We can utilize the machine learning algorithm to classify the severity of the accident based on the weather and road condition information.

The goal of accident prediction is to provide a measure of the risk of accidents at different points in time and space. The severity of an accident is the label used to train the model, and the proposed model can be used to identify where and when the risk of accident is significantly higher than average in order to take actions to reduce that risk. Note that the model cannot be used to predict whether an accident will occur or not. Indeed, in order to accurately predict the occurrence of an accident, additional data would be needed: the occurrence of an accident depends on many factors, including driver behaviour, that cannot be easily measured.

Data:

Data for the project has been obtained from the Seattle Department of Traffic (SDOT). The data contains records of all collision from year 2004 till present as recorded by traffic department.

Each record in the data corresponds to an accident and has a labelled classification “Severity” that is assigned based on the nature of the accident. Each Data records has 38 attributes(field) that characterizes each accident like unique key of the incident, address type of collision, the weather condition, road condition, over speeding, number of vehicles involved etc.

Below are the few attribute information of the data fields collected from SDOT:

|  |  |  |
| --- | --- | --- |
| Attribute | Data Type | Description |
| INCKEY | Long | Unique Key for the Incident |
| COLDETKEY | Long | Secondary Key for the Incident |
| ADDRTYPE | Text,12 | Collision address type:   * Alley * Block * Intersection |
| INTKEY | Double | Key Corresponds to the Intersection associated with Collision |
| LOCATION | Text,255 | Description of the general location of Collision |
| SEVERITYCODE | Text,100 | A code that correspond to the severity of collision:   * 3—fatality * 2b—serious injury * 2—injury * 1—prop damage * 0—unknown |
| SEVERITYDESC | Text | A detailed description of the severity of he collision |
| COLLISIONTYPE | Text,300 | Collision Type |
| PERSONCOUNT | Double | Total number of people involved in the collision |
| PEDCOUNT | Double | Total number of pedestrians involved in the collision |
| PEDCYLCOUNT | Double | Total number of bicycles involved in the collision |
| VEHCOUNT | Double | The number of vehicles involved in the collision |
| SERIOUSINJURIES | Double | The number of serious injuries involved in the collision |
| FATALITIES | Double | The number of fatalities in the collision |
| INCDATE | Date | The date of the incident |
| INCDTTM | Text ,30 | The date and time of incident |
| JUNCTIONTYPE | Text,300 | Category of junction at which collision took place |
| INATTENTIONIND | Text , 1 | Whether or not collision was due to inattention(Y/N) |
| UNDERINFL | Text,10 | Whether or not a driver involved was under the influence of drugs or alcohol |
| WEATHER | Text,300 | A description of weather condition during the time of collision |
| ROADCOND | Text,300 | The condition of road, during collision |
| LIGHTCOND | Text,300 | The light condition during the collision |
|  |  |  |