Capstone Project - Car accident severity (Week 1)

By: Leandro Recova

**1 – Introduction and Business problem.**

The goal of this capstone project is to provide a machine learning model that can predict the severity of car accident based on a dataset provided by the course.

Based on the information of previous accidents, weather condition, traffic jam, the model could provide the likelihood of an accident for a driver that might be driving along the road. This kind of warning would be very important so the driver can make decisions while driving in such conditions. This model could be integrated with a GPS software application that runs in the user application based on the trajectory followed by the driver.

**2 – Dataset Description**

The dataset that will be used during this capstone project was provided in the week 1 of the course. It has a list of 38 fields described as listed in the table below. By analyzing the data, the dataset will have to go through a pre-processing problem since some columns still have blanks, inconsistency of data description (E.g: Y, N, 0, 1 values in the same column).

Based on the fields presented in this dataset, we will have to go through a pre-processing of the fields, make sure we select the features that will have a impact in the model, and remove those columns that will not be necessarily for this project.

We will use linear regression model to predict the severity of car accident based on this dataset.

|  |  |  |
| --- | --- | --- |
| **Field** | **Field Description** | **Comments** |
| 1 | SEVERITYCODE | Code 1 or 2 depending on the accident. |
| 2 | X | Longitude of the location of the accident |
| 3 | Y | Latitude of the location of the accident |
| 4 | OBJECTID | Primary key of the table |
| 5 | INCKEY | Table Keys |
| 6 | COLDETKEY | Table Keys |
| 7 | REPORTNO | Police Report Number |
| 8 | STATUS | Matched or Unmatched status |
| 9 | ADDRTYPE | Alley, Roadblock, Intersection, and blanks |
| 10 | INTKEY | Key parameter - Some fields are blank |
| 11 | LOCATION | Address of the accident |
| 12 | EXCEPTRSNCODE | Key for except: NEI (Not enough information), blanks |
| 13 | EXCEPTRSNDESC | Description of the code if any |
| 14 | SEVERITYCODE | Severity code 1 or 2. |
| 15 | SEVERITYDESC | Two possible fields: Injury Collision and Property Damage Collision. |
| 16 | COLLISIONTYPE | Collision type. |
| 17 | PERSONCOUNT | Number of people inside the car. |
| 18 | PEDCOUNT | Pedestrians count |
| 19 | PEDCYLCOUNT | Pedestriancs cyclists count |
| 20 | VEHCOUNT | Number of vehicles. |
| 21 | INCDATE | Incdident date |
| 22 | INCDTTM | Incident date with time |
| 23 | JUNCTIONTYPE | Junction type |
| 24 | SDOT\_COLCODE | 69 different sdot codes |
| 25 | SDOT\_COLDESC | Sdot codes description |
| 26 | INATTENTIONIND | In attention in mind: Just a Yes comment. |
| 27 | UNDERINFL | Mixed of N, Y, 0, and 1 parameters. |
| 28 | WEATHER | Weather conditions |
| 29 | ROADCOND | Road conditions |
| 30 | LIGHTCOND | Light conditions |
| 31 | PEDROWNOTGRNT | Pedestrian not granted ROW indicator |
| 32 | SDOTCOLNUM | Sdot column number |
| 33 | SPEEDING | Mix of blanks and Y. |
| 34 | ST\_COLCODE | Street code |
| 35 | ST\_COLDESC | Street code description |
| 36 | SEGLANEKEY | Code |
| 37 | CROSSWALKKEY | Cross walk key |
| 38 | HITPARKEDCAR | Hit parked car (Yes or No). |