

Capstone Project - Car accident severity

Introduction

For the final capstone project in the IBM course, we need to analyze the “severity” of the accident in terms of casualties, traffic delays, property damage, or any other types of accidents. The data is collected by the Seattle SPOT Traffic Management Department and provided by Coursera via the link

(<https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>). The data set is updated weekly, from 2004 to present. It contains information such as severity code, address type, location, collision type, weather, road conditions, speeding, etc.

We can begin to solve this complex problem by studying the data of past accidents, and gain insight into the factors that affect the severity of the accident, and then use machine learning technology to create a model based on its initial similarity to predict the severity of future the historical record of the accident. Moreover, there are two main beneficiaries of this model: 1. city planner, who can use the model to inform their road planning and traffic strategies; 2. emergency service providers, who can use the model to forecast the severity of the accident based on the information provided when the accident is reported, so that resources can be optimally allocated throughout the city.

Data

Our predictor or target variable will be “SEVERITYCODE”, because it is used to measure the severity of accidents from 0 to 3. The attributes used to measure the severity of the accident are “WEATHER”, “ROADCODE”, “LIGHTCODE”.

A code that corresponds to the severity of the collision:

- 3 - fatality
- 2b - serious injury
- 2 - injury
- 1 - prop damage
- 0 - unknown

Other important variables include:

- LOCATION: Description of the general location of the collision.
- SEVERITYDESC: A detailed description of the severity of the collision.
- INCDATE: The date of the incident.
- INCDTTM: The date and time of the incident.
- JUNCTIONTYPE: Category of junction at which collision took place.
- INATTENTIONIND: Whether or not collision was due to inattention.
- UNDERINFL: Whether or not a driver involved was under the influence of drugs or alcohol.
- WEATHER: A description of the weather conditions during the time of the

collision.

- ROADCOND: The condition of the road during the collision.
- LIGHTCOND: The light conditions during the collision.
- PEDROWNOTGRNT: Whether or not the pedestrian right of way was not granted.
- SDOTCOLNUM: A number given to the collision by SDOT.
- SPEEDING: Whether or not speeding was a factor in the collision.
- SEGLANEKEY: A key for the lane segment in which the collision occurred.
- CROSSWALKKEY: A key for the crosswalk at which the collision occurred.
- HITPARKEDCAR: Whether or not the collision involved hitting a parked car.