Predicting the Severity of Road Collisions

Rachit Kumar

Introduction

Purpose and Business Case

- Car accidents cause traffic disruptions, property damage, and risk injury.
- If drivers can predict a higher risk of severe collisions, they can choose to drive more safely.
- Machine learning can be used to predict the likelihood of a serious accident based on current weather, road, and lighting conditions.

Data

Source and Description

- Collision data is sourced from the Seattle Police Department (2004 - Present)
- Attributes include the geographic coordinates of the location of the collision, the address of the collision along with address type, the severity of the results of the collision, the type of collision, each of the number of people, pedestrians, cyclists, and motorists involved in the collision, the date and time of the collision, the weather conditions, the road conditions, the road lighting conditions, whether or not the driver was speeding, whether a parked car was involved, and a description of the collision according to the collision code.

Methodology

Machine Learning Model

- The main target variable is the severity code
  1: Property damage only
  2: Injury
- Three independent variables with values coded from 0 to 8
  - Weather conditions
  - Road conditions
  - Lighting conditions
- Machine learning model used: Logistic Regression

Results and Discussion

- Logistic Regression results in an accuracy of 68%
  This is better than a blind guess.

- Possible causes of inaccuracy:

  - data do not include collisions that didn’t occur

  - sometimes harsher road conditions result in fewer severe collisions
    drivers driving more safely?

- Suggested Improvement: gather data on the approximate number of vehicles on the roads under all given conditions. Then normalise the collision data.


Drive Safe!