

Introduction

With an increase of people on the road, there has been an increase in the number of accidents, drastically impacting traffic flow as well as the safety of drivers. Over the years, the Seattle Police Department has collected data on kinds of collisions in an attempt to determine prevention measures. Collisions caused by factors such as speeding, being under the influence of drugs, and inattention can be reduced by harshly penalizing such behaviors. However, collisions due to unpredictable factors such as the weather conditions and the road conditions are more difficult to prevent. Therefore, a model needs to be developed to predict possible collisions so that police department can alert the drivers of potential collisions so they can be prepared.

The main audience for the project is the Seattle Police Department, Emergency Medical Services (EMS), as well as drivers. The goal of the developed model is to enable the police department and the EMS to predict the severity of the collisions based on the conditions present and deploy the necessary protocols. Additionally, the results of the deployment of the model can be used to alert drivers of the severity of a potential collision so that the drivers can plan accordingly beforehand.

Data

The data of the collisions was provided by the Seattle Police Department and recorded by Traffic Records, found on the Seattle Geo Data open source website. The dataset chosen for the project consists of 40 columns, and 221,266 observations or entries. The columns in the dataset include metadata such as the location of the collision, weather conditions, road conditions, and the time of the day the collision occurred. The target variable for the model generated will be the severity of the collisions provided by the “SEVERITYCODE” column. The column contains five severity codes for varying severity levels of the collisions:

- 3: fatality of parties involved
- 2b: serious injury of parties involved
- 2: injury of parties involved
- 1: property damage of parties involved
- 0: unsubstantial collision

Attributes such as the weather conditions, road conditions, and the time of day of the collision will be primarily used to predict the target variable.