Car Accident Severity Prediction in Seattle

- V. Sai Sudheer

1. Introduction

1.1 Background

Car Collision occurs when a vehicle collides with another vehicle, pedestrian, animal, road debris, or other stationary obstruction, such as a tree, pole or building. Traffic collisions often result in injury, disability, death, and property damage as well as financial costs to both society and the individuals involved. They also impact economy with increased commuting times, increased delivery times of products and costs, and pollution, due to the massive number of cars waiting for the road to be cleared, or heavily slowed down.

A number of factors contribute to the risk of collisions, including vehicle design, speed of operation, road design, road environment, driving skills, impairment due to alcohol or drugs, and behavior, notably distracted driving, speeding and street racing.

1.2 Problem

Building a model to predict the severity of car accident using traffic accident data in Seattle city of United States. This can help commuters, health workers, Administration of Seattle, professional drivers or logistic planners to reduce the personal and/or business impact of car accidents**.**

1.3 Target Audience

This model can help commuters, professional drivers, or logistic planners to reduce the personal and/or business impact of car accidents.

Seattle Government: Accident-prone areas can have Interventions such as speed breakers, Traffic signs, establish new Police check-posts for checking drunken driving, etc.. can help reduce accidents

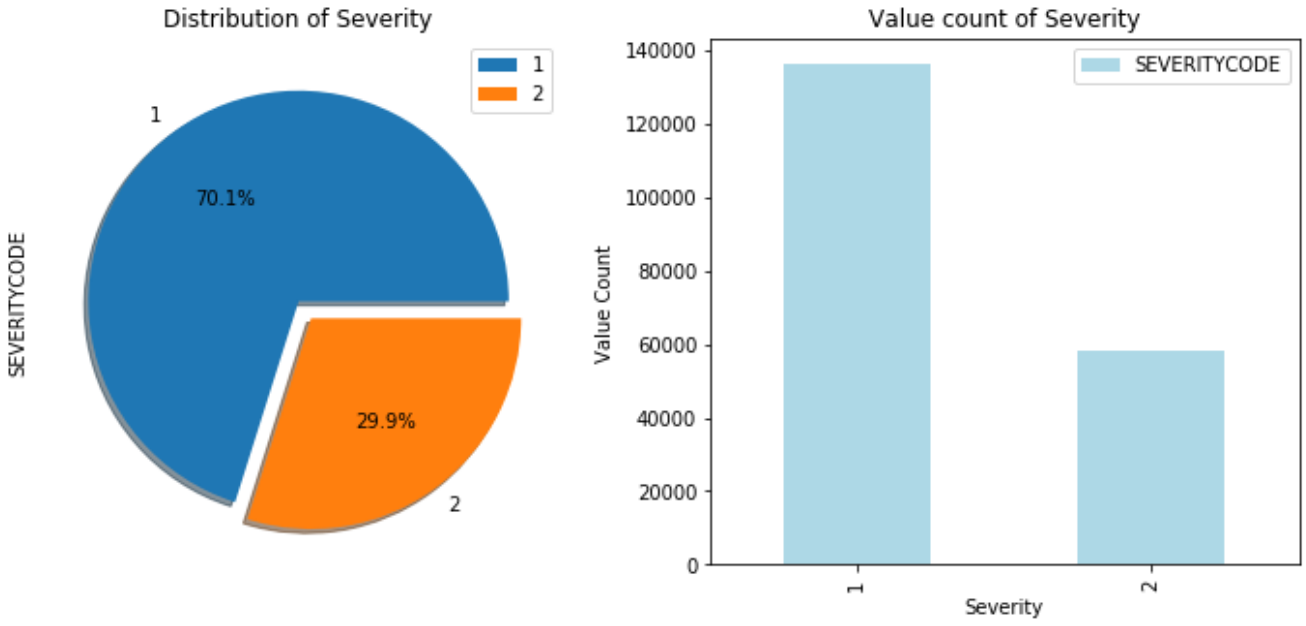
Car Owners: Owners staying in the Areas where parked cars are prone to get hit by other vehicles should concentrate on parking spots and can pay more insurance in order to decrease the loss.

Health care and Emergency services in Seattle: On predicting the severity of accidents, they can take necessary actions and can potentially save lives.

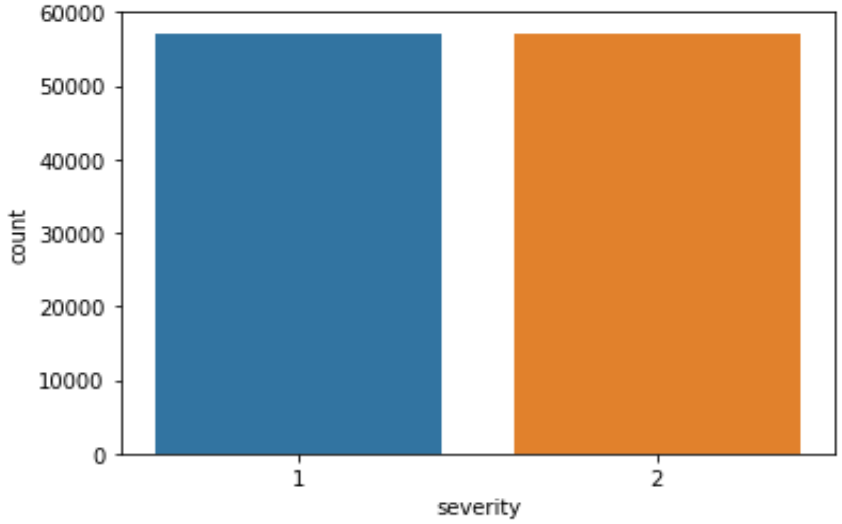
2. Data requirements

2.1

The traffic accidents data is obtained for the Seattle city in this link. The metadata description of the columns of the data is available in this link. The data period is from January 1, 2004 to May20, 2020. The column 'SEVERITYCODE' is target label that the model should predict. It is notated as 1and 2which represent property damage and injury respectively. The distribution of the severity level of the data is presented as follow. About 70.1% of accidents were with level 1 and 29.9% were with level2.



Note that the distribution between classes is not even, so under-sampling the data will help to avoid biased classification model.



After doing the under-sampling, both the classes have equal count and can potentially decrease bias.

The column INCDATE is converted into date object in the dataset.

2.2 Feature Selection

Out of 38 features, selected only 17 necessary features that are related to the problem statement. The selected features are listed below.

1. ADDRTYPE: Collision address type (Alley, Block, Intersection)
2. PERSONCOUNT: The total number of people involved in the collision
3. PEDCOUNT: The number of pedestrians involved in the accident
4. PEDCYLCOUNT: The number of bicycles involved.
5. VEHCOUNT: Number of vehicles involved.
6. UNDERINFL: Whether or not driver involved was under influence of drugs
7. WEATHER: Weather condition at the time of accident
8. ROADCOND: Road condition
9. LIGHTCOND: Light condition
10. PEDROWNOTGRNT: Whether pedestrian right of way was not granted
11. SPEEDING: Whether or not speeding was a factor of collision
12. HITPARKEDCAR: Whether collision involved hitting a parked car
13. Date: Date of accident
14. Year: Year of collision
15. Weekday: day of the week accident happened
16. Hour: Time of accident
17. Month: Month of accident