Data

To solve the problem of determining the severity of potential collisions, an effective dataset is required that has previous recorded incidents and their labeled severity along with the factors involved with that specific incident. Specifically, for this report, the dataset “Collisions – All Years” from SDOT Traffic Management Division will be used. This is a record of reported collisions in Seattle from 2004 to present day (September 2020).

The important attribute of this dataset is the labelled severity of each incident which has the following possibilities:

0 – Unknown

1 – Property Damage

2 – Injury

2b – Serious Injury

3 – Fatality

Initially, it is known that this rating system will be challenging to use due to the mix of alphanumeric characters, therefore this rating system will need to be modified to a scale of 1 to 4 and replacing the ‘2b’ label with 3, the ‘3’ label with 4, and eliminating the ‘0’ label as the unknown severity incidents do not contribute to the main solution.

There are many attributes in this dataset to help predict the severity label, and with minor adjustments in formatting, they will become effective predictors. The list of attributes used for this problem are noted below:

* ADDRTYPE (Collision address type)
* INCDTTM (The date and time of the incident)
* JUNCTIONTYPE (Category of junction)
* UNDERINFL (If the driver was under the influence or not)
* WEATHER (Description of the weather at the time of collision)
* ROADCOND (Description of the road conditions)
* LIGHTCOND (Description of the lighting conditions)
* PEDROWNOTGRNT (Whether or not the pedestrian right of was granted)
* SPEEDING (If speeding was a factor)
* ST\_COLCODE (Code relating to the description of the collision)
* HITPARKEDCAR (Whether or not the collision involved hitting a parked car)

For the attributes that are text descriptions (e.g. WEATHER), key words are isolated and used as the attribute for the prediction (e.g. “Rain”). Other formatting and data manipulation also helps other attributes work effectively with the model development.