# Machine Learning Model

## Data Section

To build this machine learning algorithm for predicting the severity of a accident; the dataset to be used to train the algorithm is as provided by SDOT Traffic Management Division, Traffic Records Group. This includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present.

The input variables or feature set would be the below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.** | **Attribute** | **Data type, length** | **Description** |
| 1 | JUNCTIONTYPE | Text, 300 | Category of junction at which collision took place |
| 2 | ROADCOND | Text, 300 | The condition of the road during the collision. |
| 3 | LIGHTCOND | Text, 300 | The light conditions during the collision. |
| 4 | WEATHER | Text, 300 | A description of the weather conditions during the time of the collision. |
| 5 | PERSONCOUNT | Double | The number of pedestrians involved in the collision. |
| 6 | VEHCOUNT | Double | The number of vehicles involved in the collision |
| 7 | SPEEDING | Text, 1 | Whether or not speeding was a factor in the collision. (Y/N) |

The labeled data would be SEVERITYCODE (Text,100) which is a code that corresponds to the severity of the collision:

• 3—fatality

• 2b—serious injury

• 2—injury

• 1—prop damage

• 0—unknown