The original dataset for this project was downloaded from the Applied Data Science Capstone Project from the following link : https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv along with its metadata pdf. I selected the most relevant attributes for use as features for the prediction of accident severity.

The features of the dataset resulting are listed below:

SEVERITYCODE – A code that corresponds to the severity of the collision:

•3—fatality • 2b—serious injury • 2—injury • 1—prop damage • 0—unknown

COLLISIONTYPE – collision type

OBJECTID – ESRI unique identifier

ADDRTYPE – Collision address type: • Alley • Block • Intersection

PERSONCOUNT – The total number of people involved in the collision

PEDCOUNT – The number of pedestrians involved in the collision. This is entered by the state.

PEDCYLCOUNT – The number of bicycles involved in the collision. This is entered by the state.

VEHCOUNT – The number of vehicles involved in the collision. This is entered by the state.

INJURIES – The number of total injuries in the collision. This is entered by the state.

SERIOUSINJURIES – The number of serious injuries in the collision. This is entered by the state.

FATALITIES – The number of fatalities in the collision. This is entered by the state.

JUNCTIONTYPE – Category of junction at which collision took place

INATTENTIONIND – Whether or not collision was due to inattention. (Y/N)

UNDERINFL – Whether or not a driver involved was under the influence of drugs or alcohol.

WEATHER – A description of the weather conditions during the time of the collision.

ROADCOND – The condition of the road during the collision.

LIGHTCOND – The light conditions during the collision.

PEDROWNOTGRNT – Whether or not the pedestrian right of way was not granted. (Y/N)

SPEEDING – Whether or not speeding was a factor in the collision. (Y/N)

HITPARKEDCAR – Whether or not the collision involved hitting a parked car. (Y/N)