Capstone Project: Car Accident Severity

1. Data Section.

- The description of the data is discussed and how it will be used to solve the problem raised.

1.1. Description of data

The Accident data (provided by seattle.gov: Link) is used to predict the Severity of an accident given certain features (Metadata). The data is for Accidents occurring in the city of Seattle from 2004 to 2020.

Label = y = SEVERITYCODE

Total Number of features: 37

Features selected (X):

Feature	Description	Reason for Selecting
ADDRTYPE	Collision at Alley, Block, Intersection	Gives the likelihood of collision at these places
PERSONCOUNT	Number of people involved in the collision	Gives an indication of severity
PEDCOUNT	Number of pedestrians involved in the accident	Gives an indication of severity
PEDCYLCOUNT	Number of cyclists involved in the accident	Gives an indication of severity
VEHCOUNT	Number of vehicles involved in the accident	Gives an indication of severity
INCDTTM	The date and time of the incident	Time of accident: midnight/ day time
INATTENTIONIND	Whether the person was not paying attention	Not paying attention can result in accident
UNDERINFL	Whether the person was driving under influence	DUI can cause accidents
WEATHER	Weather conditions	Bad weather can cause accidents
ROADCOND	Road conditions	Wet roads can cause skidding
LIGHTCOND	Light conditions	Light conditions affect visibility
PEDROWNOTGRNT	Pedestrian right of way was granted or not	
SPEEDING	Whether speeding or not	Speeding causes accidents
COLLISIONTYPE	Collision Type	Type of collision gives severity of accident
HITPARKEDCAR	Whether or not the collision involved hitting a parked car.	Hitting a parked car causes property damage

Features dropped:

Feature	Description	Reason for Dropping
X	Latitude	Can't be modelled in classification
Y	Longitude	Can't be modelled in classification
OBJECTID	ESRI unique identifier	ID not relevant
INCKEY	Secondary key for the incident	ID not relevant
COLDETKEY	Identifying key	ID not relevant
LOCATION	Description of Location	ADDRTYPE captures this
REPORTNO	Report Number	ID not relevant
STATUS	Matched/Unmatched	ID not relevant
INTKEY	Intersection key for collision	ID not relevant
EXCEPTRSNCODE	Blank	No data
EXCEPTRSNDESC	Blank	No data
SEVERITYCODE	Label	Label to be predicted
SEVERITYDESC	Description of Severity	Label to be predicted
INCDATE	The date of the incident.	INCDTTM captures this
SDOT_COLCODE	Collision code	Collision type captures this
SDOT_COLDESC	A description of the collision corresponding to the collision code.	Collision type captures this
SDOTCOLNUM	A number given to the collision by SDOT.	Collision type captures this
SEGLANEKEY	A key for the lane segment in which the collision occurred.	ID not relevant
CROSSWALKKEY	A key for the crosswalk at which the collision occurred.	ID not relevant

Features after Feature Engineering:

SL No.	Feature	Description	Reason for Selecting
1	ADDRTYPE	Collision at Alley, Block, Intersection	Gives the likelihood of collision at these places
2	PERSONCOUNT	Number of people involved in the collision	Gives an indication of severity
3	PEDCOUNT	Number of pedestrians involved in the accident	Gives an indication of severity
4	PEDCYLCOUNT	Number of cyclists involved in the accident	Gives an indication of severity
5	VEHCOUNT	Number of vehicles involved in the accident	Gives an indication of severity
6	INATTENTIONIND	Whether the person was not paying attention	Not paying attention can result in accident
7	UNDERINFL	Whether the person was driving under influence	DUI can cause accidents
8	WEATHER	Weather conditions	Bad weather can cause accidents
9	ROADCOND	Road conditions	Wet roads can cause skidding
10	LIGHTCOND	Light conditions	Light conditions affect visibility
11	PEDROWNOTGRNT	Pedestrian right of way was granted or not	
12	SPEEDING	Whether speeding or not	Speeding causes accidents

13	COLLISIONTYPE	Collision Type	Type of collision gives severity of accident
14	HITPARKEDCAR	Whether or not the collision involved hitting a parked car.	Hitting a parked car causes property damage
15	Year	Year of accident	Did one year have a lot of accidents
16	Month	Month of Accident	Does month affect number of accidents
17	Day	Day of accident	Day of month
18	Hour	Time of accident	Are accidents caused majorly at night
19	Weekday	What day of the week accident happened	Are accidents caused more on certain days of the week

1.2. Exploratory Data Analysis:

- Plotting factors on the map are used to get density of areas where accidents were caused by the features in question:
    1. Speeding:
    1. Under Influence (DUI)
    2. Inattention
    3. Hitting a parked car

- The following classifiers are used to get the prediction whether given certain attributes (features), the severity of the accident (label)
    1. K Nearest Neighbours
    2. Logistic Regression
    3. Decision Tree Classifier
    4. XGBoost Classifier
    5. Random Forest Classifier
    6. Support Vector Machine