Predicting the Severity of Car Accident

Sumit Kumar
October 20, ‘20

Introduction

1. Road Accidents are very common now
2. Given Circumstances
    a. High Traffic
    b. Weather, Light and Road Conditions
    c. Driver's condition while driving
3. Problem Statement:

" To predict the Severity of a road accident, given different attributes of the situation, like coordinates, number of vehicles, road condition, light condition, drug usage, alcohol usage during driving, not paying attention. "

4. Stakeholders
    a. Traffic Authority
    b. Woke Individuals

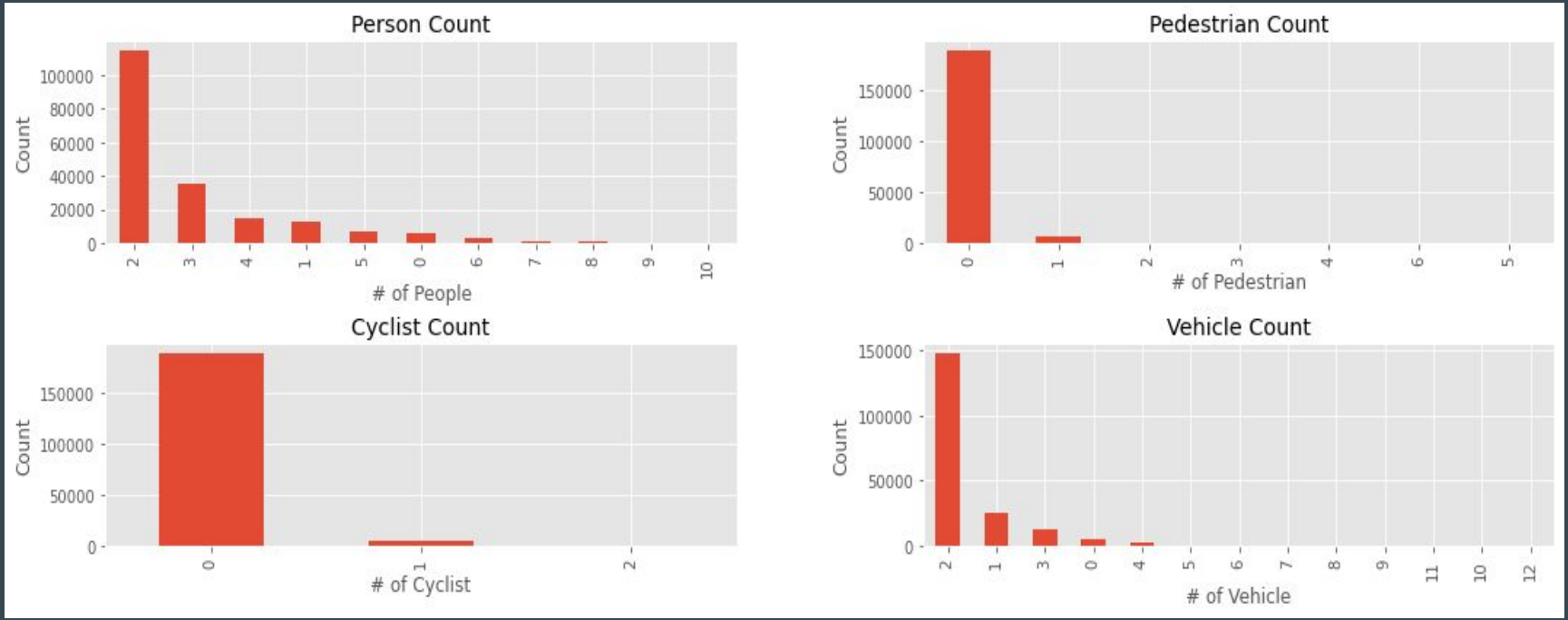
Data

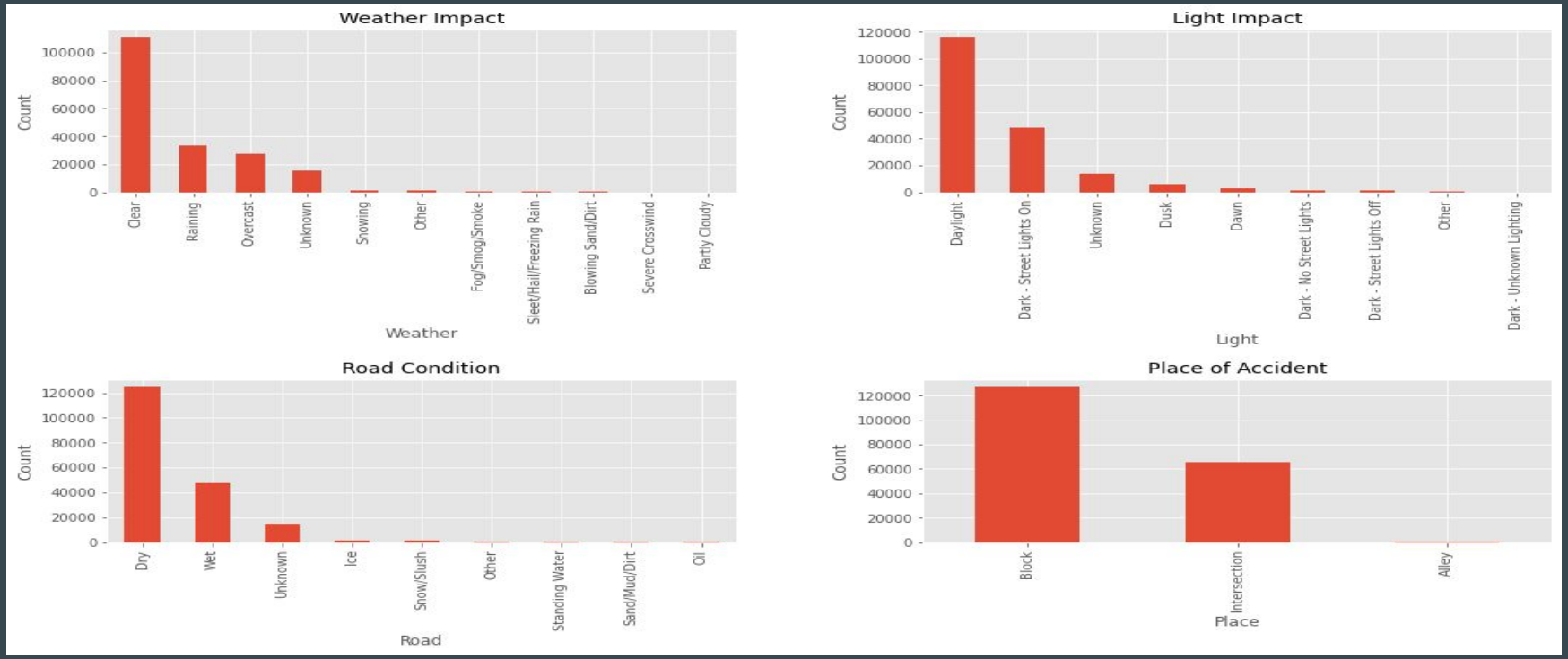
1. **Feature Selection**

['SEVERITYCODE', 'ADDRTYPE', 'INTKEY', 'COLLISIONTYPE', 'PERSONCOUNT', 'PEDCOUNT', 'PEDCYLCOUNT', 'VEHCOUNT', 'JUNCTIONTYPE', 'INATTENTIONIND', 'UNDERINFL', 'WEATHER', 'ROADCOND', 'LIGHTCOND', 'SPEEDING', 'HITPARKEDCAR']

2. **Data Cleansing and Preprocessing**
    a. **One hot encoding**
    b. **Categorical Values**

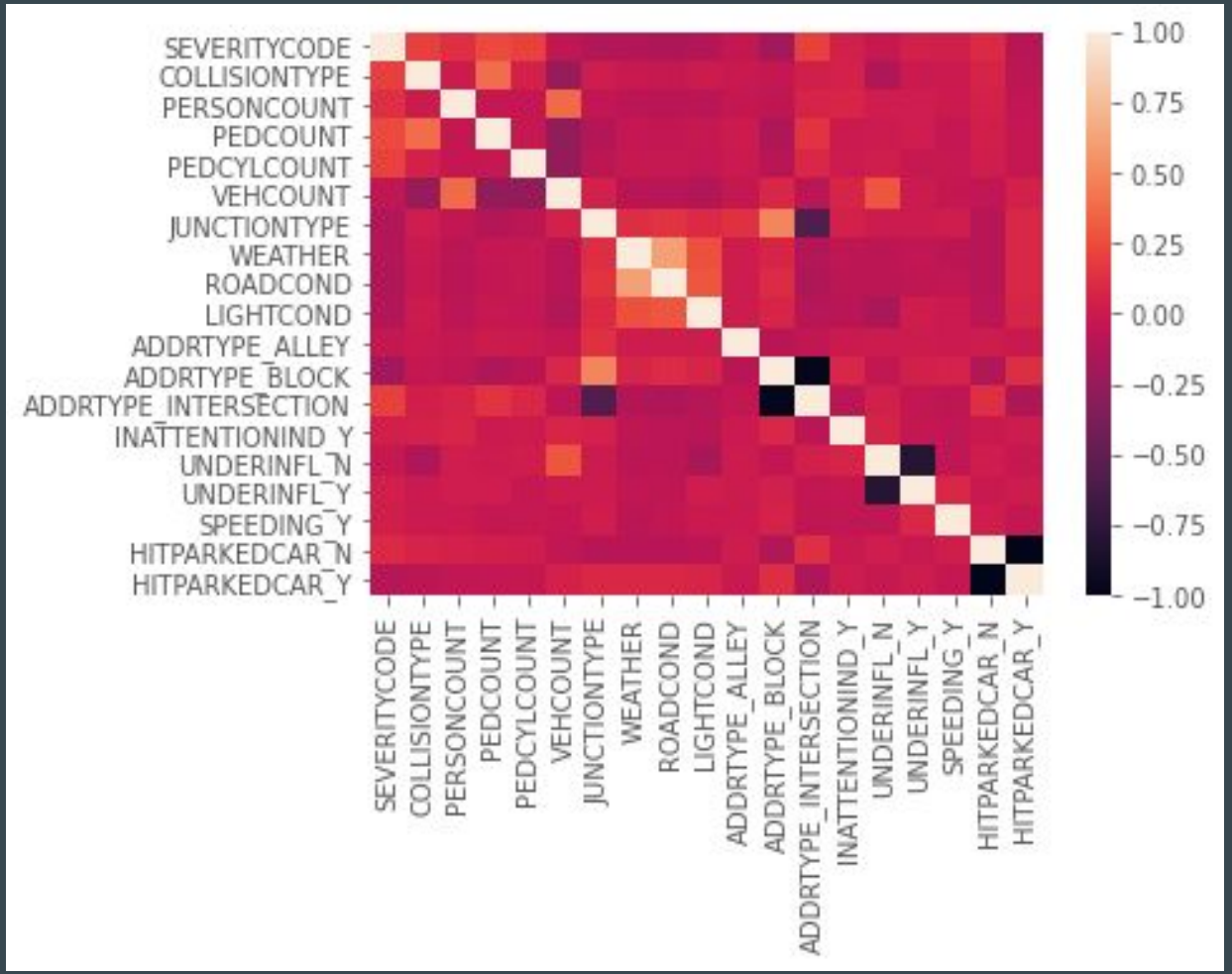
3. **Data interpretations and Visualization**


Person Count
Count
100000
80000
60000
40000
20000
0
2
3
4
1
5
0
6
7
8
9
10
# of People
Pedestrian Count
Count
150000
100000
50000
0
0
1
2
3
4
6
5
# of Pedestrian
Cyclist Count
Count
150000
100000
50000
0
0
1
2
# of Cyclist
Vehicle Count
Count
150000
100000
50000
0
2
1
3
0
4
5
6
7
8
9
11
10
12
# of Vehicle


Weather Impact
Count
100000
80000
60000
40000
20000
0
Clear
Raining
Overcast
Unknown
Snowing
Other
Fog/Smog/Smoke
Sleet/Hail/Freezing Rain
Blowing Sand/Dirt
Severe Crosswind
Partly Cloudy
Weather
Light Impact
Count
120000
100000
80000
60000
40000
20000
0
Daylight
Dark - Street Lights On
Unknown
Dusk
Dawn
Dark - No Street Lights
Dark - Street Lights Off
Other
Dark - Unknown Lighting
Light
Road Condition
Count
120000
100000
80000
60000
40000
20000
0
Dry
Wet
Unknown
Ice
Snow/Slush
Other
Standing Water
Sand/Mud/Dirt
Oil
Road
Place of Accident
Count
120000
100000
80000
60000
40000
20000
0
Block
Intersection
Alley
Place

Methodology

Correlation between Selected Features



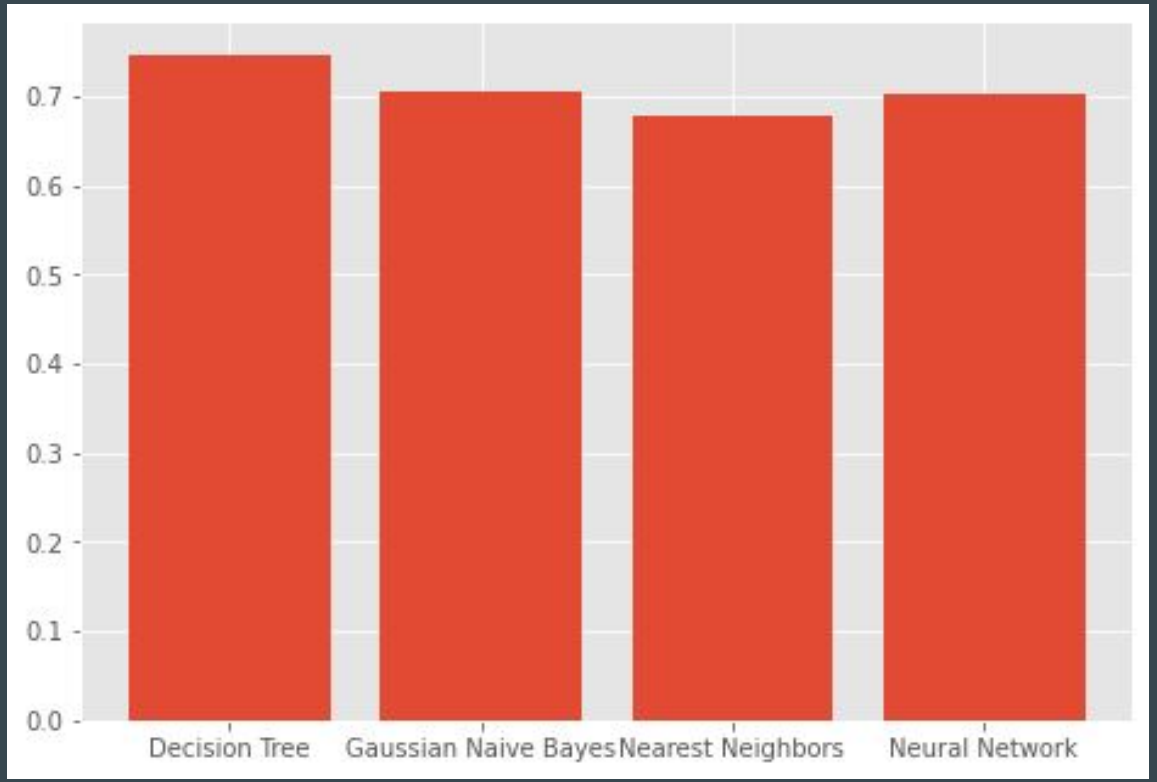
The Machine Learning Models used in the Project:

1. Decision Tree
2. Gaussian Naive Bayes
3. Nearest Neighbors
4. Neura; Networks

Result

Accuracy of Models are as follows:

1. Decision Tree: 0.747
2. Gaussian naive Bayes: 0.704
3. Nearest Neighbors: 0.697
4. Neural Network: 0.703



Conclusion and Follow Ups

1. The best accuracy was 74.7%
2. More Training is required
3. A lt of scenarios covered in data, need more extensive data
4. Better efforts to collect data

Thank You