Coursera Capstone Project Report

Project to Predict Accident Collision Severity

Introduction

Introduction ...I

Background

- Accidents in traffic lead to associated fatalities and economic losses every year and thus is an area of primary concern from loss prevention point of view
- Road traffic injuries are estimated to be the eighth leading cause of death globally and are predicted to become the seventh leading cause of death by 2030
- The proactive approach includes a collision prevention approach, like, preventing a potential unsafe road conditions from occurring in the first place

Introduction ...II

Problem

- An accurate prediction of the severity of the accidents can be performed by analysing a significant range of factors like:
  - Weather conditions
  - Road conditions
  - Traffic Jams
  - Travel Time
- Modeling accident severity prediction and improving the model are critical to the effective performance of road traffic systems for improved safety

Interest

- Insights from the model could allow Government bodies to allocate their resources more effectively in advance and create warning systems / faster reactive processes

Data Acquisition and Cleaning

Data Acquisition and Cleaning ...I

Data Acquisition

- This project is based on kaggle datasets that contain an extended descriptions of different aspect of the accidents
- The most relevant and useful data is considered for analysis

```
In [4]: df_accident_data = pd.read_csv(project.get_file("Accident_Data.csv"))

In [5]: df_accident_data.head()
Out[5]:
```

	Collision_Ref_No	Policing_Area	Collision_Severity	Weekday_of_Collision	Day_of_Collision	Month_of_Collision	Hour_of_Collision	Carriagev
0	3518	CREA	3	MON	4	8	14.0	13
1	10557	BELC	3	SAT	8	8	17.0	11
2	5002	LISB	3	WED	5	11	17.0	1
3	11714	BELC	3	SUN	18	10	16.0	12
4	12416	MIDU	3	MON	23	11	9.0	13

Basic Data Exploration

```
In [6]: df_accident_data.shape
Out[6]: (8849, 17)
```

Data Acquisition and Cleaning ...II

Data Cleaning

- The most relevant and useful data is considered for analysis and encoding used for training and test data

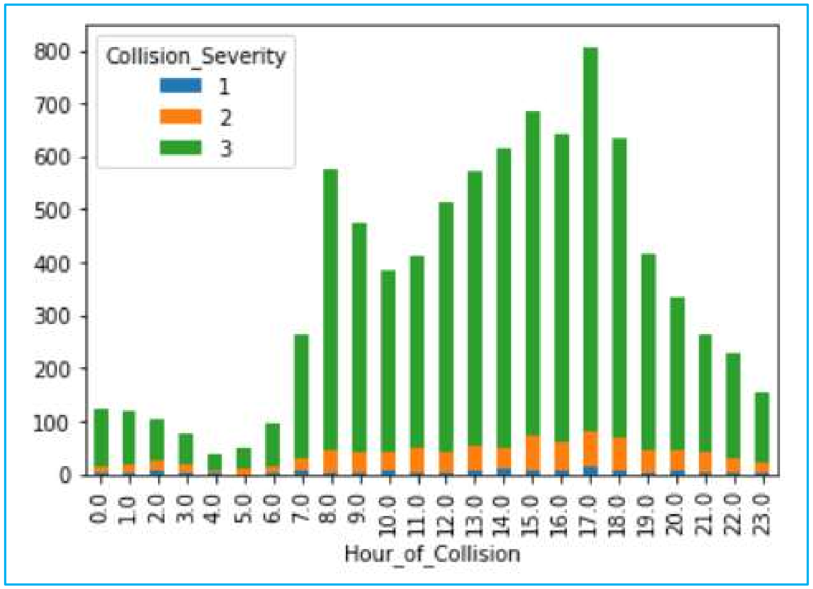
```
In [14]: data_train = pd.read_csv(project.get_file("Accident_Data.csv"))
         data_test = pd.read_csv(project.get_file("Accident_TestData.csv"))

In [15]: le = LabelEncoder()
         le.fit(data_train['Weekday_of_Collision'])
         data_train['Weekday_of_Collision'] = le.transform(data_train['Weekday_of_Collision'])
         le.fit(data_train['Policing_Area'].astype(str))
         data_train['Policing_Area'] = le.transform(data_train['Policing_Area'].astype(str))
         data_train = data_train.fillna(data_train.median())

In [16]: le.fit(data_test['Collision_Severity'])
         data_test['Collision_Severity'] = le.transform(data_test['Collision_Severity'])
         le.fit(data_test['Weekday_of_Collision'])
         data_test['Weekday_of_Collision'] = le.transform(data_test['Weekday_of_Collision'])
         le.fit(data_test['Policing_Area'].astype(str))
         data_test['Policing_Area'] = le.transform(data_test['Policing_Area'].astype(str))
         data_test = data_test.fillna(data_test.median())

In [17]: labels_train = data_train["Collision_Severity"]
         data_train_x = data_train.drop(["Collision_Severity", "Collision_Ref_No"], axis=1)
         data_test_x = data_test.drop(["Collision_Severity", "Collision_Ref_No"], axis=1)
         data_train_x.shape

Out[17]: (8849, 15)
```

Exploratory Data Analysis

Exploratory Data Analysis ...I

Collision Severity v/s. Hour of Collision

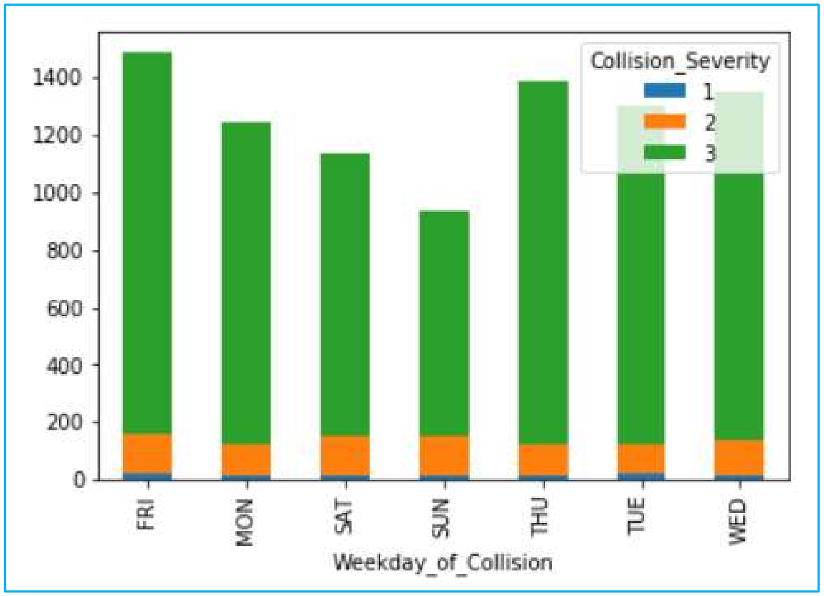
- Based on the available data, plotting the collision severity against the hour of collision



Exploratory Data Analysis ...II

Collision Severity v/s. Weekday of Collision

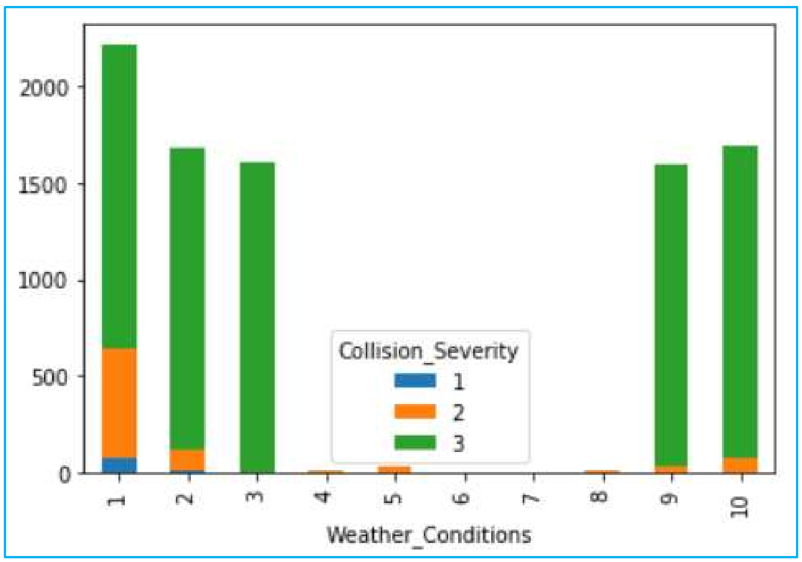
- Based on the available data, plotting the collision severity against the weekday of collision



Exploratory Data Analysis ...III

Collision Severity v/s. Weather Conditions

- Based on the available data, plotting the collision severity against the weather conditions

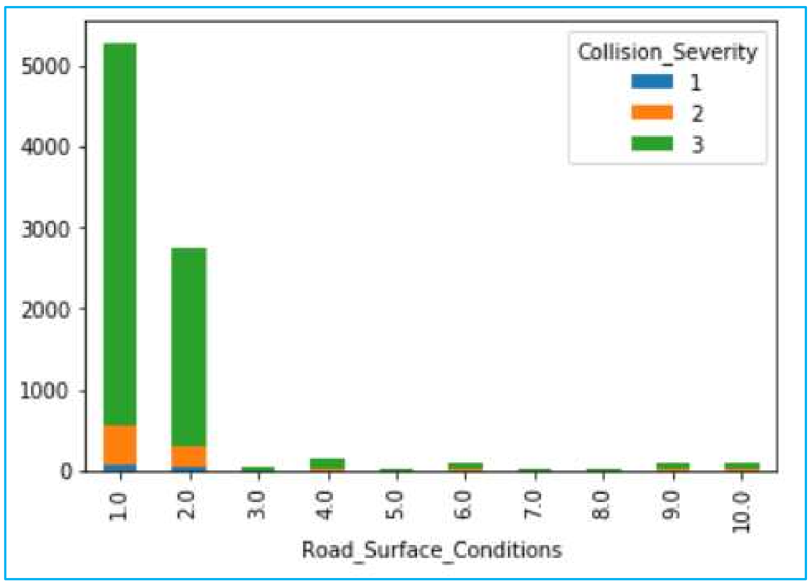



```
1  :  Fine without high winds
2  :  Raining without high winds
3  :  Snowing without high winds
4  :  Fine with high winds
5  :  Raining with high winds
6  :  Snowing with high winds
7  :  Fog or mist - if hazard
8  :  Strong sun (glaring)
9  :  Other
10 :  Unknown
```

Exploratory Data Analysis ...IV

Collision Severity v/s. Road Surface Conditions

- Based on the available data, plotting the collision severity against the road surface conditions




```
1  :  Dry
2  :  Wet / damp
3  :  Snow
4  :  Frost / ice
5  :  Flood
6  :  Oil
7  :  Mud
8  :  Leaves
9  :  Slippery (after dry spell)
10 :  Other
```

Modeling

Modeling ...I

Applying Standard Algorithms

- Using the available dataset to train the model and then using the test dataset to evaluate the accuracy for collision severity prediction
- Different Model Algorithms used are listed below:

Model Algorithms
Decision Tree
KNN
Linear SVC
Logistic Regression
Naive Bayes
Perceptron
Random Forest
Stochastic Gradient Descent
Support Vector Machines
XG Boost

Modeling ...II

Performances of different Models

- Result of the evaluation of each model based on test dataset to accurately predict the collision severity

- Metric sorted as per model accuracy

Model	Score	Collision Severity Count Details
XGBoost	90.42	{3: 2166, 2: 46, 1: 1}
Random Forest	89.79	{3: 2183, 2: 29, 1: 1}
Support Vector Machines	89.47	{3: 2213}
Logistic Regression	89.47	{3: 2213}
Linear SVC	89.47	{3: 2213}
Stochastic Gradient Descent	87.75	{3: 2128, 2: 80, 1: 5}
KNN	86.99	{3: 2105, 2: 84, 1: 24}
Decision Tree	85.18	{3: 1945, 2: 222, 1: 46}
Perceptron	79.80	{3: 1891, 2: 321, 1: 1}
Naive Bayes	9.81	{1: 1982, 3: 193, 2: 38}

Conclusion

Conclusion

- As seen from the results table, XG Boost along with its advantage for execution speed has also given the most accurate results for the model when evaluated against the test data for predicting collision severity

Future Direction

Future Direction

- Data analysis can help build useful models to predict the severity of a traffic accident
- Accuracy of the model has room for improvement and various other factors like speed, travel time, weather conditions, road conditions, peak traffic window can be corelated and analysed for improved results
- These models can then be enhanced to predict potentail accident spots and time of the day along with the collision severity, which in turn can provide Authorities insight and time to develop procedures and policies for betterment of all