

Capstone Project Car Accident Analysis

Applied Data Science Capstone by IBM/Coursera

By : Mostafa Eltayeb

Table of contents

1. Introduction: Business Problem
2. Data
3. Visualization
4. Methodology
5. Analysis
6. Results and Discussion
7. Conclusion

Introduction: Business Problem

Car Accidents is one of major cause of mortality around the world ,

Sometimes, accidents occur for **a combination of reasons**, from **bad visibility** to unsafe **road design**, or other drivers lack caution.

While the causes of accidents can vary, the consequences are often the same, resulting in everything from vehicular and property damage to serious injuries.

Introduction: Business Problem

We need to **reduce** the rate of accident by **previous warning** the driver and give them more info and precautions .

With the aid of previous year's Datasets of similar accidents and weather conditions we can :

1. **predict the severity** of accidents expected and prevent them or
2. have a **quick response** .

Data

Based on definition of our problem, factors that will influence our decision are:

- data collected from previous collisions at last year/years
- current weather condition , driver , road .

2.1 Data sources

We have a dataset

name : Collisions—All Years

provided by : SPD and recorded by Traffic Records.

about : the impact of traffic

2.1 Data sources

After Investigation :
dimentions , Types , columns ,rows

Result :

Dataset contain (194673 **rows** , 38 **columns**)
also :

- 1- redundancy like 'SEVERITYCODE.1','SEVERITYCODE'
Columns
- 2- missing values
- 3- Unneeded columns in our investigation

2.2 Data cleaning

- 1- Remove redundancy : dropping column **'SEVERITYCODE.1'**
- 2- missing values : **Drop Null values**
- 3- Unneeded columns in our investigation

2.3 Feature selection

3- Select only column we need in our investigation

```
In [11]: df_acc = df[['SEVERITYCODE', 'ROADCOND', 'WEATHER', 'LIGHTCOND', 'X', 'Y', 'ADDRTYPE']]
df_acc.head()
```

```
Out[11]:
```

	SEVERITYCODE	ROADCOND	WEATHER	LIGHTCOND	X	Y	ADDRTYPE
0	2	Wet	Overcast	Daylight	-122.323148	47.703140	Intersection
1	1	Wet	Raining	Dark - Street Lights On	-122.347294	47.647172	Block
2	1	Dry	Overcast	Daylight	-122.334540	47.607871	Block
3	1	Dry	Clear	Daylight	-122.334803	47.604803	Block
4	2	Wet	Raining	Daylight	-122.306426	47.545739	Intersection

Label encoding

we need to use label encode to make dataset ready for analysis

6]:

	SEVERITYCODE	ROADCOND	WEATHER	LIGHTCOND	X	Y	ADDRTYPE	WEATHER_CAT	ROADCOND_CAT	LIGHTCOND_CAT	ADDRTYPE_CA
0	2	Wet	Overcast	Daylight	-122.323148	47.703140	Intersection	2	1	0	
1	1	Wet	Raining	Dark - Street Lights On	-122.347294	47.647172	Block	1	1	1	
2	1	Dry	Overcast	Daylight	-122.334540	47.607871	Block	2	0	0	
3	1	Dry	Clear	Daylight	-122.334803	47.604803	Block	0	0	0	
4	2	Wet	Raining	Daylight	-122.306426	47.545739	Intersection	1	1	0	
...
5556	2	Dry	Clear	Daylight	-122.359398	47.697849	Intersection	0	0	0	
5557	1	Dry	Clear	Daylight	-122.268968	47.538904	Block	0	0	0	
5558	1	Dry	Overcast	Daylight	-122.305693	47.604331	Block	2	0	0	
5559	2	Dry	Clear	Daylight	-122.301865	47.719316	Intersection	0	0	0	
5560	1	Wet	Raining	Daylight	-122.314095	47.733964	Block	1	1	0	

5279 rows × 11 columns

< >

Data Visualization

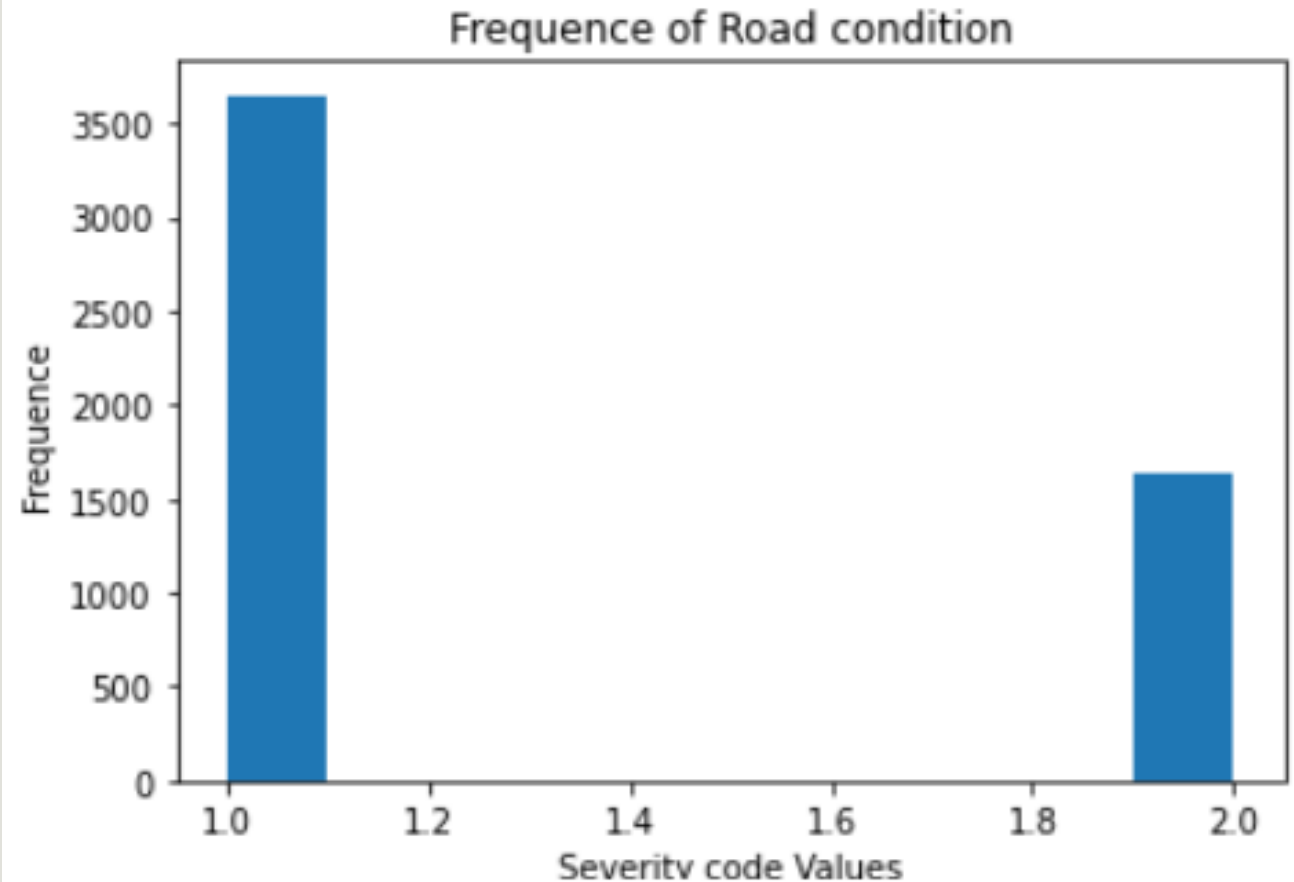
Our Target Column is

SEVERITYCODE

We notice that data is unbalanced

And need to be balanced

To prevent bias

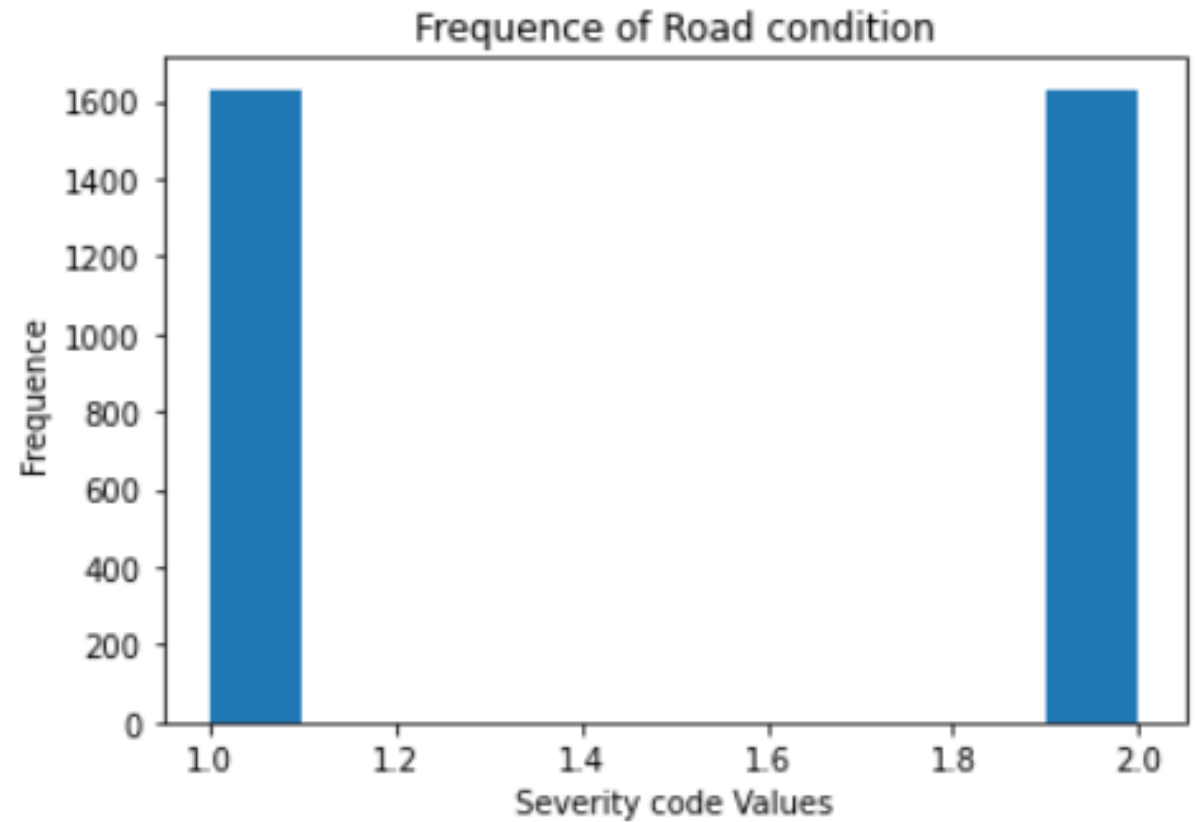


Data Visualization

Our Target Column is

SEVERITYCODE

Here dataset is **balanced**



Methodology

Classification Model

- 1- K-Nearest Neighbors**
- 2-Decision Tree**
- 3-Support Vector Machine**
- 4-Logistic Regression**

Result

```
3]:
```

	Algorithm	Jaccard	F1-score	LogLoss
0	KNN	0.672727	0.509295	NA
1	Decision Tree	0.659342	0.603268	NA
2	SVM	0.656523	0.601756	NA
3	Logistic Regression	0.659386	0.602881	24.2043

Discussion

Our analysis shows that we can predict the **Accident severity** with accuracy around **0.67** based on our dataset using classification models like decision tree , k-value , Logistic Regression

Conclusion and future directions

Purpose of this project was to analyzed the relationship between various features and Traffic Accident Severity. Using classification models to predict Traffic Accident Severity.

We Can in the future enhance the accuracy of these models using more analytics methods and methodologies.