CAR ACCIDENT SEVERITY FORECAST FOR DRIVER CAUTION AND ASSISTANCE SYSTEM

INTRODUCTION & BUSINESS PROBLEMS:

Car accidents are issues found across the globe to be severe and dangerous. Accidents might sometimes due to natural reasons or to our negligence or other reasons. A Road Traffic Accident according to Garber (2010) is a random event involving a road user that results into property damage, death or injury. Road Traffic Accidents (RTA) has become not just a National but a Global issue. It is on records that Road traffic accidents cause an estimated 13 million deaths and 20-50 million disabilities worldwide annually with 85% of injury related deaths occurring in developing countries. According to WHO (2007), RTA injuries accounted for 23% of all injury deaths worldwide. This project discusses the severity of car accidents citing City of Seattle.

Determining the causes of Road Traffic accidents Prediction of Road Traffic accidents Determining the effects and severity of Road Traffic Accidents This project is useful to the Road Safety Corporation and stakeholders interested in the welfare and safety of road users.

- Objective to predict the severity of car accidents through reliable and robust machine learning model
- Many factors such as weather, road and lighting conditions influence the severity of accidents
- Results are of particular interest to Seattle transportation department
- Beneficial to warn drivers in case of bad weather/light/Road conditions
- Help in improved Driver assistance systems and capable of handling such challenging situations during the given conditions

DATA UNDERSTANDING AND PRE-PROCESSING:

- Collision reports from Seattle (2004-2014) by the Seattle Police Department and Traffic Records department is used
- Data has multiple independent variables and 221,144 records
- Independent variables: “WEATHER”, “ROADCOND” and “LIGHTCOND”
- Dependent variable/ target: “SEVERITYCODE” (0 to 4 levels)

	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	INJURIES	SERIOUSINJURIES	FATALITIES	ADDRTYPE
0	0	0	0	0	0	0	0	0	Block
1	0	0	0	0	0	0	0	0	NaN
2	1	2	0	0	2	0	0	0	Block
3	1	2	0	0	2	0	0	0	Block
4	1	2	0	0	2	0	0	0	Block

COLLISIONTYPE	...	UNDERINFL	PEDROWNOTGRNT	SPEEDING	HITPARKEDCAR	WEATHER	ROADCOND	LIGHTCOND	YEAR	MONTH	DAY
NaN	...	NaN	NaN	NaN	N	NaN	NaN	NaN	2004	11	10
NaN	...	NaN	NaN	NaN	N	NaN	NaN	NaN	2010	04	07
Left Turn	...	N	NaN	NaN	N	Overcast	Dry	Daylight	2013	04	02
Parked Car	...	0	NaN	NaN	N	Clear	Dry	Other	2007	03	06
Rear Ended	...	0	NaN	NaN	N	Clear	Dry	Daylight	2006	05	17

- Rows that have null values are dropped
- Attributes that are not used are also dropped
- Columns are of type ‘str are encoded to ‘int’

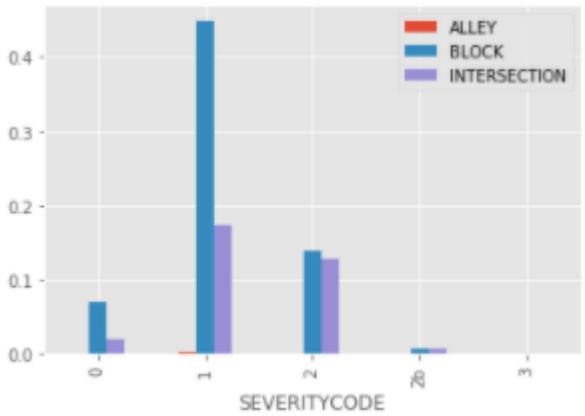
```
N      103455
0       81676
nan     26384
Y        5399
1        4230
Name: UNDERINFL, dtype: int64
```

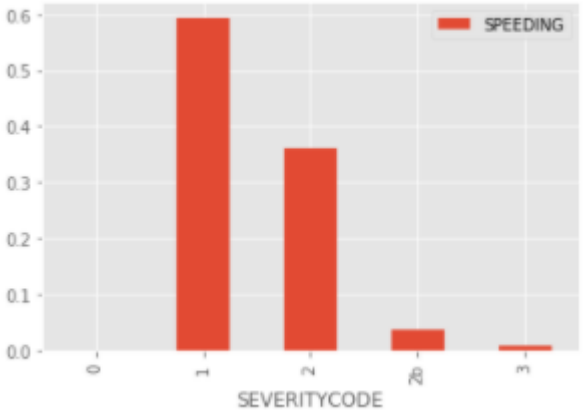
DATA VISUALIZATION:

The predictions based on the features that are not directly related with the collisions

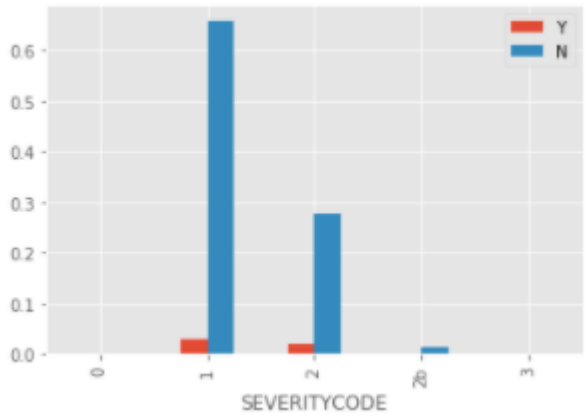
	SEVERITYCODE	ADDRTYPE	COLLISIONTYPE	INATTENTIONIND	UNDERINFL	WEATHER	ROADCOND	LIGHTCOND	PEDROWNOTGRNT	SPEEDING	HITPARKEDCAR
0	0	Block	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
2	1	Block	Left Turn	NaN	N	Overcast	Dry	Daylight	NaN	NaN	N
3	1	Block	Parked Car	NaN	0	Clear	Dry	Other	NaN	NaN	N
4	1	Block	Rear Ended	Y	0	Clear	Dry	Daylight	NaN	NaN	N

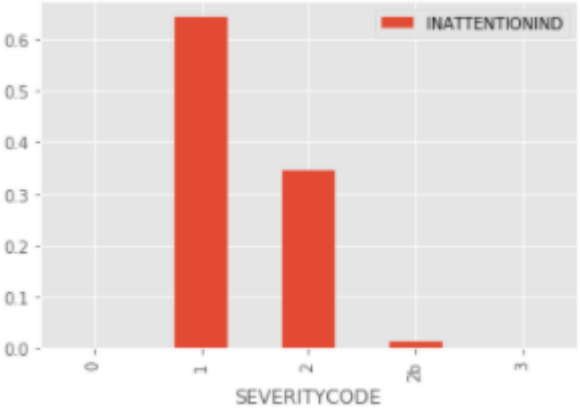
In Relation with Address type & SPEEDING



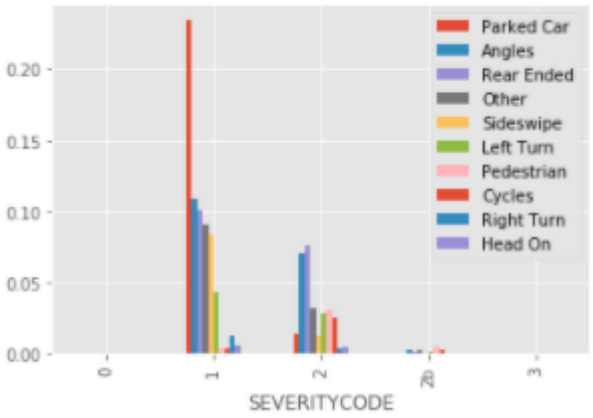


In Relation with the Under Influence Category & INATTENTION INDICATOR

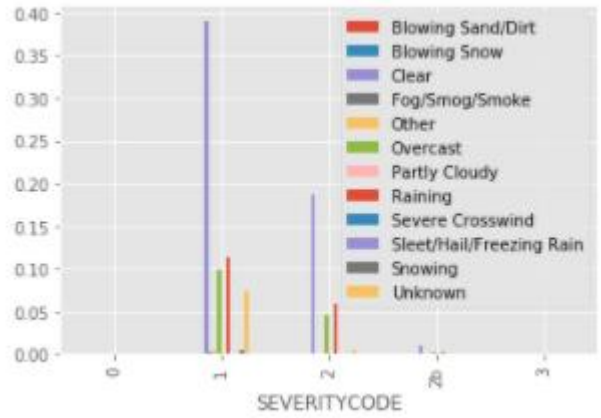


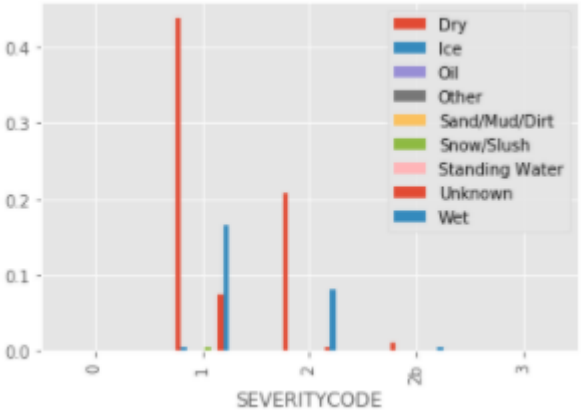


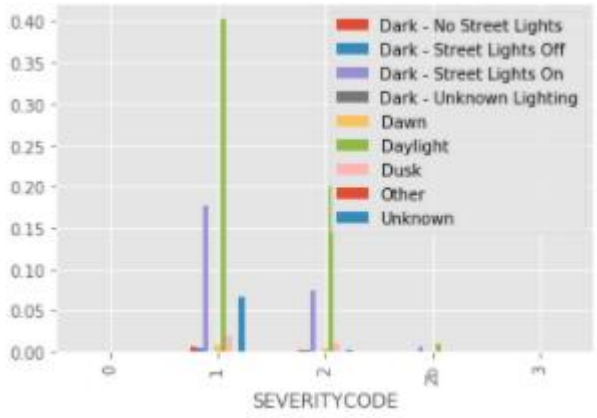
In Relation with the COLLISIONTYPE and plot given below



In Relation with the Environment:







Project Machine Learning Algorithm is used in car collisions and severity recognition.

Last Data Set:

	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	INJURIES	SERIOUSINJURIES	FATALITIES	INATTENTIONIND	UNDERINFL	...	SPEEDING	HITPARKEDCAR	ADDRTYPE	COLLISIONTYPE	WEATHER
0	0	0	0	0	0	0	0	0	NaN	NaN	...	NaN	0	Block	NaN	NaN
1	0	0	0	0	0	0	0	0	NaN	NaN	...	NaN	0	NaN	NaN	NaN
2	1	2	0	0	2	0	0	0	NaN	0.0	...	NaN	0	Block	Left Turn	Overcast
3	1	2	0	0	2	0	0	0	NaN	0.0	...	NaN	0	Block	Parked Car	Clear
4	1	2	0	0	2	0	0	0	1	0.0	...	NaN	0	Block	Rear Ended	Clear

5 rows × 21 columns

PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	INJURIES	SERIOUSINJURIES	FATALITIES	INATTENTIONIND	UNDERINFL	...	SPEEDING	HITPARKEDCAR	ADDRTYPE	COLLISIONTYPE	WEATHER	ROADCOND	LIGHTCOND	YEAR	MONTH	DAY
0	0	0	0	0	0	NaN	NaN	...	NaN	0	Block	NaN	NaN	NaN	NaN	2004	11	10
0	0	0	0	0	0	NaN	NaN	...	NaN	0	NaN	NaN	NaN	NaN	NaN	2010	04	07
0	0	2	0	0	0	NaN	0.0	...	NaN	0	Block	Left Turn	Overcast	Dry	Daylight	2013	04	02
0	0	2	0	0	0	NaN	0.0	...	NaN	0	Block	Parked Car	Clear	Dry	Other	2007	03	06
0	0	2	0	0	0	1	0.0	...	NaN	0	Block	Rear Ended	Clear	Dry	Daylight	2006	05	17

	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	INJURIES	SERIOUSINJURIES	FATALITIES	INATTENTIONIND	UNDERINFL	PEDROWNOTGRNT	SPEEDING	HITPARKEDCAR	YEAR	MONTH	DAY
0	0	0	0	0	0	0	0	0	NaN	NaN	NaN	NaN	0	2004	11	10
1	0	0	0	0	0	0	0	0	NaN	NaN	NaN	NaN	0	2010	04	07
2	1	2	0	0	2	0	0	0	NaN	0.0	NaN	NaN	0	2013	04	02
3	1	2	0	0	2	0	0	0	NaN	0.0	NaN	NaN	0	2007	03	06
4	1	2	0	0	2	0	0	0	1	0.0	NaN	NaN	0	2006	05	17

CONCLUSION:

Useful machine learning models to predict the severity of car accidents. The most of the algorithms are biased towards most frequent class. However, efficient pre-processing and corresponding imbalanced data techniques should give optimal results.