Full Report: Car Accident Severity

A. Introduction

A.1. Description & Disscusion of the Background

Road Accident is the most undesirable and unexpected thing to occur to a road user, though they happen quite often. Unfortunately, we can see a minatory rise of road accidents in United Kingdom, conspicuously highroad accidents over the past few years. It has a massive impact on society as well as in the economy of our country as there is an immense cost of fatalities and injuries. According to a recent report, annually on an average 9,000 lives have been taken by road accidents and lead to almost 28,000 injuries. This record indicates that every day, approximately 28 people were killed by road accidents and it is quite devastating. Besides this, according to WHO, the economic cost of road accidents to a developing country like us is 2-3% of GDP, which is a significant loss for a country like ours. Moreover, reducing this loss has become a great matter of concern for our country now.

A.2. Data Description

For the accurate prediction of the severity of accidents, a considerable number of traffic accident records with full information is required to train by using the proposed approaches. In this research work, the authors have collected a dataset from the Traffic Bureau that consists of total 37,885 traffic accidents record from the year 2007-2017. The entire dataset will split into two parts- Training Dataset and Test Dataset. 70% of the whole dataset has been chosen randomly by using a python library as a training data set and the remaining 30% has been used as our test dataset. We have used the 70-30 ratio for splitting dataset because of its proven accuracy.

Table 1a: Data Gathering

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	...	ROADCOND
0	2	-122.323148	47.703140	1	1307	1307	3502005	Matched	Intersection	37475.0	...	Wet
1	1	-122.347294	47.647172	2	52200	52200	2607959	Matched	Block	NaN	...	Wet
2	1	-122.334540	47.607871	3	26700	26700	1482393	Matched	Block	NaN	...	Dry
3	1	-122.334803	47.604803	4	1144	1144	3503937	Matched	Block	NaN	...	Dry
4	2	-122.306426	47.545739	5	17700	17700	1807429	Matched	Intersection	34387.0	...	Wet

Table 1b: Data Gathering

LIGHTCOND	PEDROWNOTGRNT	SDOTCOLNUM	SPEEDING	ST_COLCODE	ST_COLDESC	SEGLANEKEY	CROSSWALKKEY	HITPARKEDCAR
Daylight	NaN	NaN	NaN	10	Entering at angle	0	0	N
Dark - Street Lights On	NaN	6354039.0	NaN	11	From same direction - both going straight - bo...	0	0	N
Daylight	NaN	4323031.0	NaN	32	One parked--one moving	0	0	N
Daylight	NaN	NaN	NaN	23	From same direction - all others	0	0	N
Daylight	NaN	4028032.0	NaN	10	Entering at angle	0	0	N

B. Methodology

As a database, I used GitHub repository in my study. My master data which has the main components Severity Code, Weather Elements, Light Condition, Junction Type and Collission Type. For classification problems, matplotlib is extensively used the supervised algorithm. The primary perspective of this algorithm is predicting the value of the desired variable by learning decision rules deduced from the features of the data and create a model of that. A root node is designated for the construction of this model based on the best attribute picked by the gain approach and the sub-nodes are then generated on the basis of the decision taken in relation to the status of quality selected at each node. When each node is reduced to a single quality status, the class is determined at the end of the

node; it is called a leaf. These courses of action continue recursively until a class is defined at the end of each node.

Table 2a: Data Interpolation

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	INTKEY	SEVERITYCODE.1
count	194673.000000	189339.000000	189339.000000	194673.000000	194673.000000	194673.000000	65070.000000	194673.000000
mean	1.298901	-122.330518	47.619543	108479.364930	141091.456350	141298.811381	37558.450576	1.298901
std	0.457778	0.029976	0.056157	62649.722558	86634.402737	86986.542110	51745.990273	0.457778
min	1.000000	-122.419091	47.495573	1.000000	1001.000000	1001.000000	23807.000000	1.000000
25%	1.000000	-122.348673	47.575956	54267.000000	70383.000000	70383.000000	28667.000000	1.000000
50%	1.000000	-122.330224	47.615369	106912.000000	123363.000000	123363.000000	29973.000000	1.000000
75%	2.000000	-122.311937	47.663664	162272.000000	203319.000000	203459.000000	33973.000000	2.000000
max	2.000000	-122.238949	47.734142	219547.000000	331454.000000	332954.000000	757580.000000	2.000000

Table 2b: Data Interpolation

PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	SDOT_COLCODE	SDOTCOLNUM	SEGLANEKEY	CROSSWALKKE
194673.000000	194673.000000	194673.000000	194673.000000	194673.000000	1.149360e+05	194673.000000	1.946730e+05
2.444427	0.037139	0.028391	1.920780	13.867768	7.972521e+06	269.401114	9.782452e+03
1.345929	0.198150	0.167413	0.631047	6.868755	2.553533e+06	3315.776055	7.226926e+04
0.000000	0.000000	0.000000	0.000000	0.000000	1.007024e+06	0.000000	0.000000e+00
2.000000	0.000000	0.000000	2.000000	11.000000	6.040015e+06	0.000000	0.000000e+00
2.000000	0.000000	0.000000	2.000000	13.000000	8.023022e+06	0.000000	0.000000e+00
3.000000	0.000000	0.000000	2.000000	14.000000	1.015501e+07	0.000000	0.000000e+00
81.000000	6.000000	2.000000	12.000000	69.000000	1.307202e+07	525241.000000	5.239700e+06

Table 3a: Weather Condition

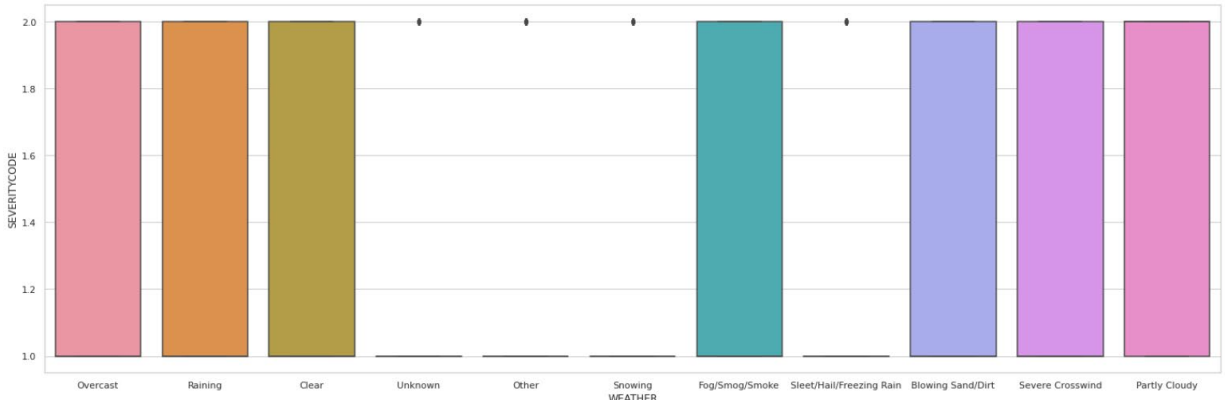


Table 3b: Weather Condition

	WEATHER
Clear	111135
Raining	33145
Overcast	27714
Unknown	15091
Snowing	907
Other	832
Fog/Smog/Smoke	569
Sleet/Hail/Freezing Rain	113
Blowing Sand/Dirt	56
Severe Crosswind	25
Partly Cloudy	5

Table 4: Road and Light Condition

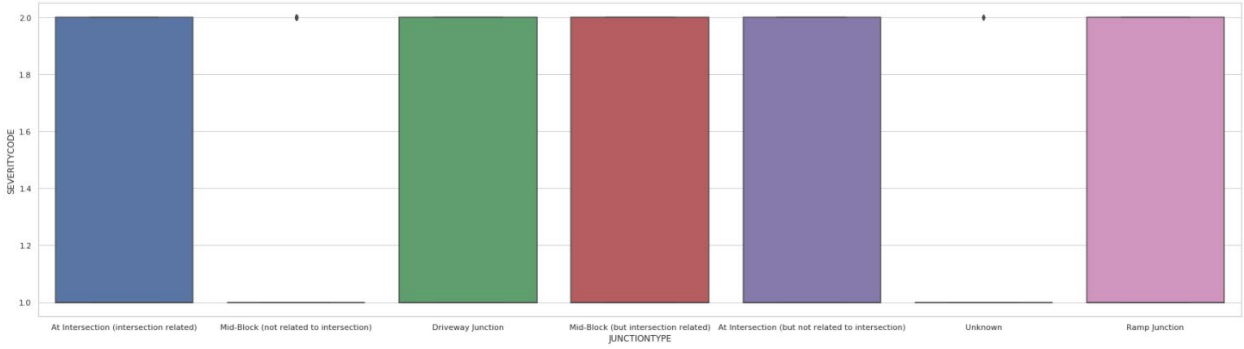
	ROADCOND
Dry	124510
Wet	47474
Unknown	15078
Ice	1209
Snow/Slush	1004
Other	132
Standing Water	115
Sand/Mud/Dirt	75
Oil	64

	LIGHTCOND
Daylight	116137
Dark - Street Lights On	48507
Unknown	13473
Dusk	5902
Dawn	2502
Dark - No Street Lights	1537
Dark - Street Lights Off	1199
Other	235
Dark - Unknown Lighting	11

Table 5a: Junction Type

	JUNCTIONTYPE
Mid-Block (not related to intersection)	89800
At Intersection (intersection related)	62810
Mid-Block (but intersection related)	22790
Driveway Junction	10671
At Intersection (but not related to intersection)	2098
Ramp Junction	166
Unknown	9

Table 5b: Junction Type



C. Results

In this research paper, I have determined the performance of each algorithm, for four accident severity classes (Fatal / Grievous /Simple Injury/ Motor Collision). By overall performance, Ada-Boost gives the best result because of its iterative classification on matplotlib.

Table 6: Combination of all Data Gathered

	WEATHER	ROADCOND	JUNCTIONTYPE	SPEEDING	ST_COLCODE	SEVERITYCODE
1	Blowing Sand/Dirt	Snow/Slush	Mid-Block (not related to intersection)	Y	50	2.0
11	Clear	Dry	At Intersection (but not related to intersection)	Y	13	2.0
18	Clear	Dry	At Intersection (intersection related)	Y	0	2.0
19	Clear	Dry	At Intersection (intersection related)	Y	1	2.0
28	Clear	Dry	At Intersection (intersection related)	Y	25	2.0
33	Clear	Dry	At Intersection (intersection related)	Y	30	2.0
35	Clear	Dry	At Intersection (intersection related)	Y	45	2.0
39	Clear	Dry	At Intersection (intersection related)	Y	73	2.0
41	Clear	Dry	At Intersection (intersection related)	Y	0	2.0
42	Clear	Dry	At Intersection (intersection related)	Y	1	2.0
49	Clear	Dry	At Intersection (intersection related)	Y	2	2.0
50	Clear	Dry	At Intersection (intersection related)	Y	21	2.0
53	Clear	Dry	At Intersection (intersection related)	Y	24	2.0
68	Clear	Dry	Driveway Junction	Y	13	2.0
70	Clear	Dry	Driveway Junction	Y	16	2.0
75	Clear	Dry	Driveway Junction	Y	45	2.0
87	Clear	Dry	Driveway Junction	Y	45	2.0
89	Clear	Dry	Driveway Junction	Y	52	2.0
96	Clear	Dry	Mid-Block (but intersection related)	Y	15	2.0
98	Clear	Dry	Mid-Block (but intersection related)	Y	21	2.0

D. Discussion

We observe that most of the accidents in our dataset are Fatal and value for the other three classes is very low. For that reason, in second experiment, we merge Grievous, Simple Injury, Motor Collision these three accident severity classes into one class. Therefore, we have attained the performances of the proposed approaches for two accident severity classes (Fatal / Grievous). In this experiment, we have noticed that the accuracy of interpolated data get increased and remain the same. But it is also mentionable that, the performance is much better than the previous experiment as precision and F1 score increased here in a noticeable way. Besides this, we did experiment with the features in our dataset and have tried to find out their effect on a traffic accident. Statistically I have found that based on the condition of some features the number of accidents gets increased. It's a significant noticeable thing for making proper steps to decrease the number of accidents.

F. Conclusion

As a result, people are turning to big cities to start a business or work. From the above table, we can see that servere car accidents occurs frequently under clear and dry condition at intersection. Besides, speeding is also an important factor leading the accident happen.