OPIM 5604-B13
Predictive Modeling Group Project:
Severity Level of Vehicle Accidents in UK

——Bowen Cai, Mengdi Zheng, Ming Zhu, Fangyuan Li

Outline

- Motivation & Problem Definition
- Analysis
- Main Findings
- Implications & Actionable Plans
- Q&A


秦EV


30
30
National
Windscreens


Zona Leste
Centro
Centro


850 DHQ


SCOTT
DARE
STUNT


Auto
Insurance

Data

- Name: UK Car Accidents 2005-2015
- Source: Kaggle.com

VARIABLES		
Accident Circumstances	**Vehicle**	**Casualty**
Accident Index	Accident Index	Accident Index
Police Force	Vehicle Reference	Vehicle Reference
Accident Severity	Vehicle Type	Casualty Reference
Number of Vehicles	Towing and Articulation	Casualty Class
Number of Casualties	Vehicle Manoeuvre	Sex of Casualty
Date (DD/MM/YYYY)	Vehicle Location-Restricted Lane	Age of Casualty
Day of Week	Junction Location	Age Band of Casualty
Time (HH:MM)	Skidding and Overturning	Casualty Severity
Location Easting OSGR (Null if not known)	Hit Object in Carriageway	Pedestrian Location
Location Northing OSGR (Null if not known)	Vehicle Leaving Carriageway	Pedestrian Movement
Longitude (Null if not known)	Hit Object off Carriageway	Car Passenger
Latitude (Null if not known)	1st Point of Impact	Bus or Coach Passenger
Local Authority (District)	Was Vehicle Left Hand Drive?	Pedestrian Road Maintenance Worker (From 2011)
Local Authority (Highway Authority - ONS code)	Journey Purpose of Driver	Casualty Type
1st Road Class	Sex of Driver	Casualty IMD Decile
1st Road Number	Age of Driver	Casualty Home Area Type
Road Type	Age Band of Driver	
Speed limit	Engine Capacity	
Junction Detail	Vehicle Propulsion Code	
Junction Control	Age of Vehicle (manufacture)	
2nd Road Class	Driver IMD Decile	
2nd Road Number	Driver Home Area Type	
Pedestrian Crossing-Human Control		
Pedestrian Crossing-Physical Facilities		
Light Conditions		
Weather Conditions		
Road Surface Conditions		
Special Conditions at Site		
Carriageway Hazards		
Urban or Rural Area		
Did Police Officer Attend Scene of Accident		
Lower Super Ouput Area of Accident_Location (England & Wales only)		

1,074,826 Rows

Topic/Problem Definition

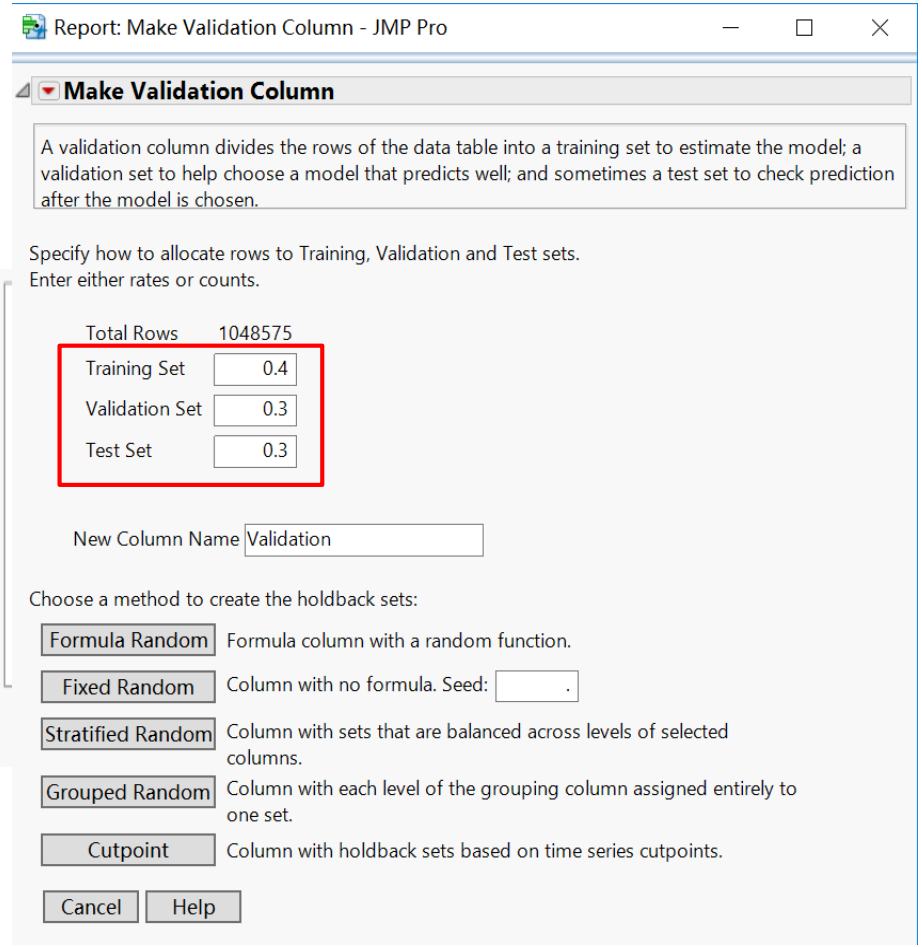
- Classify the severity level of vehicle accidents in UK
- Find significant contributing factors

Functions of Project

- Help local police to deal with car accidents easily
- Provide valuable information for insurance companies and drive safety education

Analysis-SEMMA

- Sample
- Explore
- Modify
- Model
- Assess



Learning Curve

Explore

Longitude <dbl>	Latitude <dbl>	Accident_Severity <int>	Number_of_Vehicles <int>	Number_of_Casualties <int>	Month <int>	Time <int>	Day_of_Week <int>	Road_Type <int>	Speed_Limit <int>
-0.191170	51.48910	2	1	1	1	17	3	6	30
-0.211708	51.52007	3	1	1	1	17	4	3	30
-0.206458	51.52530	3	2	1	1	0	5	6	30
-0.173862	51.48244	3	1	1	1	10	6	6	30
-0.156618	51.49575	3	1	1	1	21	2	6	30
-0.203238	51.51554	3	2	1	1	12	3	6	30

Junction_Control <int>	Light_Conditions <int>	Urban_or_Rural_Area <int>	Casualty_Class <int>	Age_of_Casualty <int>	Casualty_Severity <int>	Vehicle_Type <int>
NA	1	1	3	37	2	9
2	4	1	2	37	3	11
NA	4	1	1	62	3	11
NA	1	1	3	30	3	9
NA	NA	1	1	49	3	9

Vehicle_Manoeuvre <int>	Journey_Purpose_of_Driver <int>	Sex_of_Driver <int>	Age_of_Driver <int>	Age_of_Vehicle <int>	High_Winds_or_Not <int>
NA	NA	2	74	NA	2
4	1	1	42	3	1
17	1	1	35	5	1
2	NA	1	62	6	1
NA	NA	2	49	4	1
NA	NA	1	49	10	2

Modify: Missing Values

Summary Statistics

15 Columns Clear Select Distribution

Columns	N	N Missing
Longitude	1048464	111
Latitude	1048464	111
Road_Type	1041309	7266
Junction_Control	694704	353871
Light_Conditions	1039203	9372
Urban_or_Rural_Area	1048432	143
Age_of_Casualty	1020955	27620
Vehicle_Type	1040686	7889
Vehicle_Manoeuvre	548412	500163
Journey_Purpose_of_Driver	286593	761982
Sex_of_Driver	989690	58885
Age_of_Driver	931742	116833
Age_of_Vehicle	710616	337959
High_Winds_or_Not	1000697	47878
Road_Surface_Conditions	1047386	1189

Summary of Missing Values for Predictors

Impute:

Continuous-Mean, Nominal-Mode

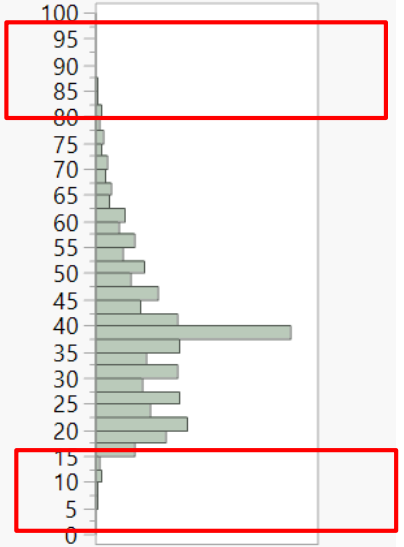
Delete:

Longitude, Latitude, Age_of_Casualty, Age_of_Driver

Modify: Outliers

Distributions

Age_of_Driver

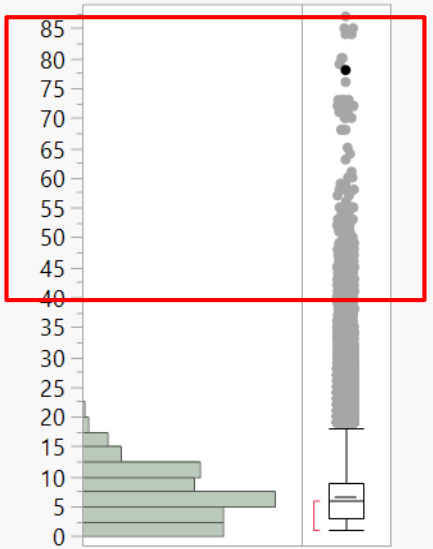


Quantiles

100.0%	maximum	99
99.5%		83
97.5%		73
90.0%		58
75.0%	quartile	46
50.0%	median	37.735505108
25.0%	quartile	26
10.0%		20
2.5%		17
0.5%		12
0.0%	minimum	1

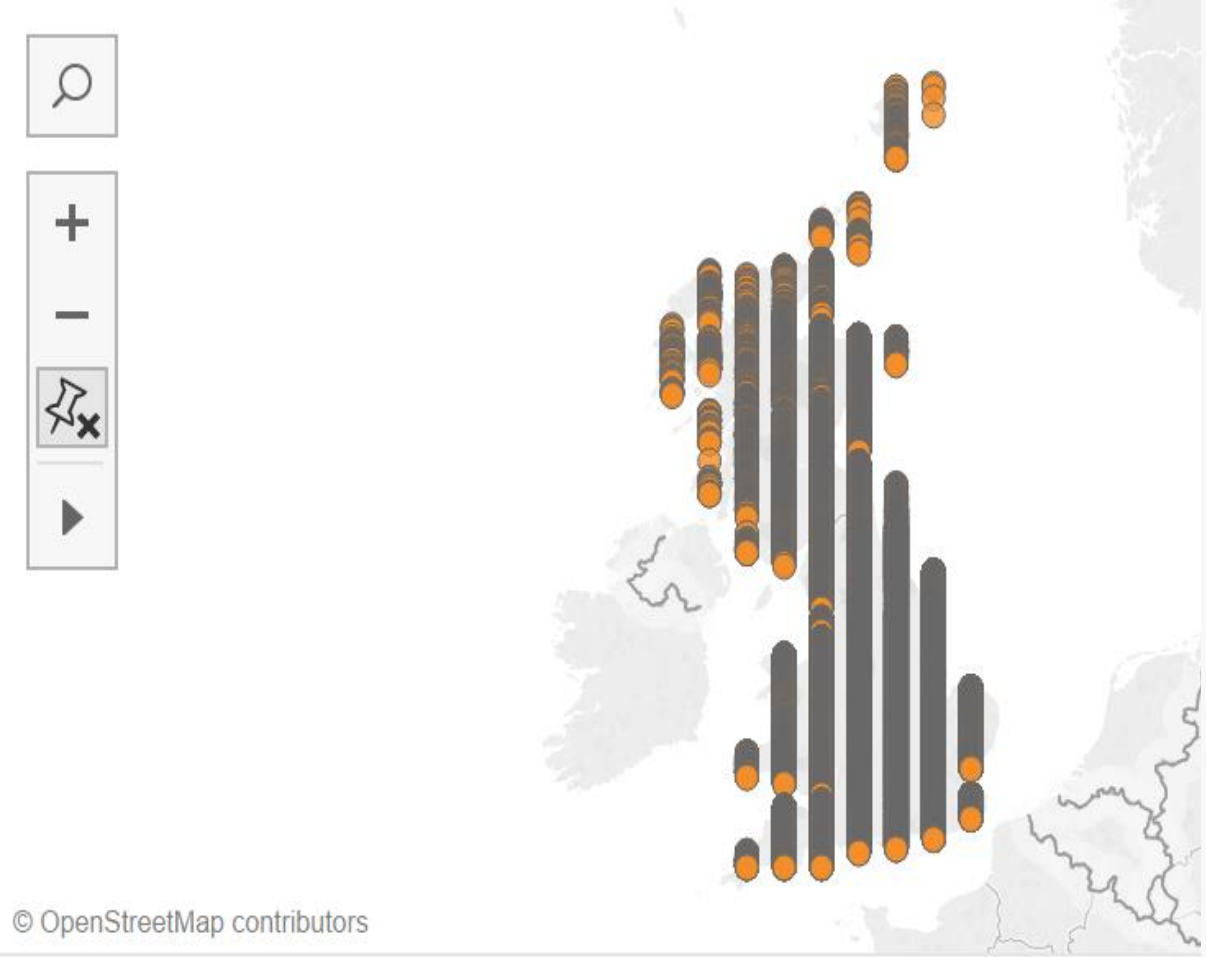
Distributions

Age_of_Vehicle



Quantiles

100.0%	maximum	87
99.5%		21
97.5%		16
90.0%		12
75.0%	quartile	9
50.0%	median	6
25.0%	quartile	3
10.0%		2
2.5%		1
0.5%		1
0.0%	minimum	1



Iterative Clustering

Cluster Comparison

Method	NCluster	CCC	Best
K-Means Clustering	8	-195.26	Optimal CCC

Columns Scaled Individually

Control Panel

K Means NCluster=8

Columns Scaled Individually

Cluster Summary

Cluster	Count
1	65779
2	131792
3	304247
4	163661
5	26296
6	21635
7	150739
8	8919

Step	Criterion
30	0

Cluster Means

Cluster	Longitude	Latitude
1	-3.7003709	51.1149119
2	-2.5920095	53.4244232
3	0.00567909	51.6116446
4	-1.7381692	51.916908
5	-4.3290821	55.9168474
6	-3.0572735	55.5504537
7	-1.2834841	53.7972571
8	-2.7721951	57.2594912

Visualization for Longitude and Latitude in Tableau

Modify: Variable Transformation

1	Variables	Explaination
2	Accident_Severity	Severity of the accident (1:Serious/2:Slight)
3	Location	Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5, Cluster 6, Cluster 7, Cluster 8
4	Number_of_Vehicles	Number of vehicles involved in the accident
5	Number_of_Casualties	Number of casualties involved in the accident
6	Month	Spring, Summer, Fall, Winter
7	Time	Daytime or Night
8	Day_of_Week	Weekday or Weekend
9	Road_Type	Type of road (1:Roundabout/2:One way street/3:Dual carriageway/ 6:Single carriageway/7:Slip road/9:Unknown/12:One way street/-1:Others)
10	Speed_Limit	Speed limit of the road (10, 15, 20, 30, 40, 50, 60, 70)
11	Junction_Control	Junction condition (0:Not at junction or within 20 metres/1:Authorised person/ 2:Auto traffic signal/3:Stop sign/4:Give way or uncontrolled/-1:Others)

Columns View Selector

Select Columns

87 Columns

- Accident_Severity
- Number_of_Vehicles
- Number_of_Casualties
- Urban?
- Casualty_Severity
- Sex_of_Driver = male?
- High_Winds_or_Not
- Validation
- Winter?
- Fall?
- Summer?
- Spring?
- Weekday?
- Weekend?
- Daytime?
- Night?
- Road type = slip road?
- Road type = single carriageway?
- Road type = dual carriageway?
- Road type = one way street?
- Road type = roundabout?
- Speed limit = 70?
- Speed limit = 60?
- Speed limit = 50?
- Speed limit = 20?
- Speed limit = 15?
- Speed limit = 10?
- Speed limit = 40?
- Speed limit = 30?
- Junction control = others?

Columns View Selector

Select Columns

87 Columns

- Junction control = others?
- Junction control = uncontrolled?
- Junction control = controlled?
- Light conditions = no lighting?
- Light conditions = lights unlit?
- Light conditions = lights lit?
- Light conditions = daylight?
- Casualty class = pedestrain?
- Casualty class = passenger?
- Casualty class = driver?
- Casualty = aged?
- Casualty = bet middle and aged?
- Casualty = middle-aged?
- Casualty = teenager?
- Casualty = youth?
- Casualty = long lived?
- Casualty = baby?
- Vehicle type = bus?
- Vehicle type = car?
- Vehicle type = motorcycle?
- Vehicle type = cycle?
- Journey purpose of drivers = school?
- Journey purpose of driver = work?
- Road surface condition = flood?
- Road surface condition = ice?
- Road surface condition = snow?
- Road surface condition = wet?
- Road surface condition = dry?
- Vehicle manoeuvre = go ahead?
- Vehicle manoeuvre = overtake?

Columns View Selector

Select Columns

87 Columns

- Vehicle manoeuvre = go ahead?
- Vehicle manoeuvre = overtake?
- Vehicle manoeuvre = change line?
- Vehicle manoeuvre = turn right?
- Vehicle manoeuvre = turn left?
- Vehicle manoeuvre = static?
- Vehicle manoeuvre = reverse?
- Age of driver = 70s?
- Age of driver = 60s?
- Age of driver = 50s?
- Age of driver = 40s?
- Age of driver = 30?
- Age of driver = 20s?
- Age of driver = 10s?
- Location = Cluster 8?
- Location = Cluster 7?
- Location = Cluster 6?
- Location = Cluster 5?
- Location = Cluster 4?
- Location = Cluster 3?
- Location = Cluster 2?
- Location = Cluster 1?
- Age of vehicle < 6
- Age of vehicle = (6, 10)?
- Age of vehicle = (11, 15)?
- Age of vehicle = (16, 20)?
- Age of vehicle = (21, 25)?
- Age of vehicle = (26, 30)?
- Age of vehicle = (31, 35)?
- Age of vehicle = (36, 40)?

Parameter Collinearity Effect

Contingency Table

Junction control = uncontrolled?

Road surface condition = dry?	0	1	Total
0 (Count / Total % / Col % / Row %)	58199 / 5.81 / 31.81 / 18.19	261686 / 26.14 / 31.98 / 81.81	319885 / 31.95
1	124746 / 12.46 / 68.19 / 18.31	556504 / 55.59 / 68.02 / 81.69	681250 / 68.05
Total	182945 / 18.27	818190 / 81.73	1001135

Tests

N	DF	-LogLike	RSquare (U)
1001135	1	1.0087073	0.0000

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	2.017	0.1555
Pearson	2.016	0.1556

Contingency Table

Summer?

Road surface condition = flood?	0	1	Total
0 (Count / Total % / Col % / Row %)	746070 / 74.52 / 99.86 / 74.63	253677 / 25.34 / 99.85 / 25.37	999747 / 99.86
1	1016 / 0.10 / 0.14 / 73.20	372 / 0.04 / 0.15 / 26.80	1388 / 0.14
Total	747086 / 74.62	254049 / 25.38	1001135

Tests

N	DF	-LogLike	RSquare (U)
1001135	1	0.73638459	0.0000

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	1.473	0.2249
Pearson	1.491	0.2221

Contingency Table

Winter?

Road type = slip road?	0	1	Total
0 (Count / Total % / Col % / Row %)	757033 / 75.62 / 98.95 / 76.42	233609 / 23.33 / 98.96 / 23.58	990642 / 98.95
1	8036 / 0.80 / 1.05 / 76.58	2457 / 0.25 / 1.04 / 23.42	10493 / 1.05
Total	765069 / 76.42	236066 / 23.58	1001135

Tests

N	DF	-LogLike	RSquare (U)
1001135	1	0.07948316	0.0000

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.159	0.6901
Pearson	0.159	0.6903

Contingency Table

Road type = single carriageway?

Road surface condition = wet?	0	1	Total
0 (Count / Total % / Col % / Row %)	176929 / 17.67 / 71.11 / 24.83	535524 / 53.49 / 71.18 / 75.17	712453 / 71.16
1	71883 / 7.18 / 28.89 / 24.90	216799 / 21.66 / 28.82 / 75.10	288682 / 28.84
Total	248812 / 24.85	752323 / 75.15	1001135

Tests

N	DF	-LogLike	RSquare (U)
1001135	1	0.24412090	0.0000

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.488	0.4847
Pearson	0.488	0.4847

-Methodology Adopted: Chi-Square

- Road Surface = Dry? vs Junction Control = Uncontrolled?
- Road Surface Condition = Good? vs Summer?
- Road Type= Slip Road? vs Winter?
- Road Surface Condition = Wet? vs Road Type = Single Carriageway?

Parameter Collinearity Effect

Findings:

- Keep: “road_surface_condition = dry, wet, flood”, “Summer”, “Winter”
- Delete: “Junction control = uncontrolled?/controlled’’

Confusion Rates: Training (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.26%	11.74%
2	77.42%	22.58%

Confusion Rates: Validation (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.56%	11.44%
2	77.42%	22.58%

Confusion Rates: Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.49%	11.51%
2	77.48%	22.52%

Parameter Collinearity Effect

Conclusion:

- It is best to delete dimension = Junction control
- Mitigate collinearity in logistic model makes difference

Confusion Rates: Training (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.30%	11.70%
2	77.71%	22.29%

19037 rows have been excluded.

Confusion Rates: Validation (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.49%	11.51%
2	77.80%	22.20%

14219 rows have been excluded.

Confusion Rates: Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.56%	11.44%
2	77.73%	22.27%

14184 rows have been excluded.

Assess Selected Models

Three Primary Clarification:

- Assess based on the permutation-missing-value-dataset as defaulted.
- Use test set for each model with the same rate (Training =0.4, Validation=0.3, Test=0.3).
- The misclassification rates displayed are all based on model after the cut-off adjustment.

Stepwise: Best Subset Variable Selection

Test

Actual Accident_Severity	Decision Count 1	Decision Count 2
1	31853	6263
2	150521	73284

Specified Profit Matrix

Actual	Decision 1	Decision 2
1	0	-1
2	-0.111	0

Actual Accident_Severity	Decision Rate 1	Decision Rate 2
1	0.836	0.164
2	0.673	0.327

Misclassification Rate
0.5986

Test

Actual Accident_Severity	Decision Count 1	Decision Count 2
1	31534	6582
2	148095	75710

Specified Profit Matrix

Actual	Decision 1	Decision 2
1	0	-1
2	-0.111	0

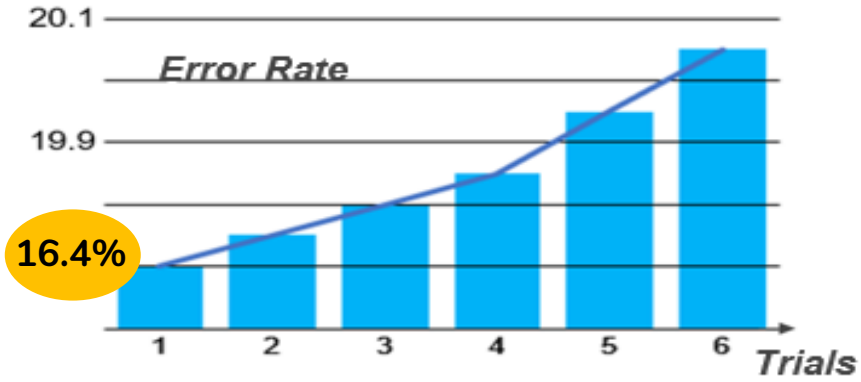
Actual Accident_Severity	Decision Rate 1	Decision Rate 2
1	0.827	0.173
2	0.662	0.338

Model Performance of Boosted Tree: Forward Selection & Mixed Selection

Learning Rate & Tuning Parameter Analysis

Tuning

Increasing learning rate and number of trees, splits in boosted tree.



Confusion Rates:
Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.56%	11.44%
2	77.73%	22.27%

11.44%

Logistic Regression

Test

Actual	Decision Count	
Accident_Severity	1	2
1	31853	6263
2	150521	73284

Actual	Decision Rate	
Accident_Severity	1	2
1	0.836	0.164
2	0.673	0.327

Boosted Tree

Stepwise: Best Subset Variable Selection

Confusion Rates: Training (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.12%	10.88%
2	78.89%	21.11%

19037 rows have been excluded.

Confusion Rates: Validation (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.31%	10.69%
2	79.00%	21.00%

14219 rows have been excluded.

Confusion Rates: Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.35%	10.65%
2	78.90%	21.10%

14184 rows have been excluded.

47440 rows have been excluded.

Confusion Rates: Training (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.67%	11.33%
2	78.23%	21.77%

19037 rows have been excluded.

Confusion Rates: Validation (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.88%	11.12%
2	78.33%	21.67%

14219 rows have been excluded.

Confusion Rates: Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	88.94%	11.06%
2	78.24%	21.76%

14184 rows have been excluded.

Model Performance of Logistic: Forward Selection & Mixed Selection

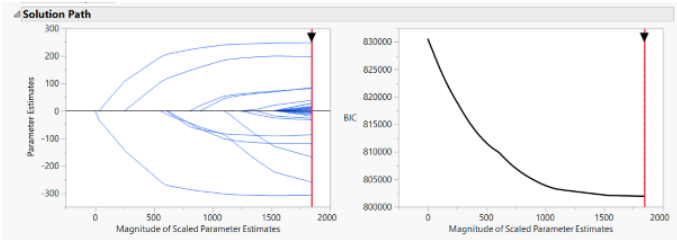
Tuning Parameter Using LASSO Penalization with BIC

Effect Tests

Source	Nparm	DF	Wald ChiSquare	Prob > ChiSquare	
Number_of_Vehicles	1	1	5617.6101	<.0001*	
Number_of_Casualties	1	1	6678.7595	<.0001*	
Urban?	1	1	4101.0169	<.0001*	
Road type = single carriageway?	1	1	1461.6308	<.0001*	
Road surface condition = dry?	1	1	813.36299	<.0001*	
Location = Cluster 5?	1	1	585.27194	<.0001*	
Daytime?	1	1	570.66235	<.0001*	
Light conditions = daylight?	1	1	395.00557	<.0001*	
Road surface condition = wet?	1	1	357.40407	<.0001*	
Location = Cluster 8?	1	1	171.82767	<.0001*	
Road type = slip road?	1	1	116.45295	<.0001*	
Location = Cluster 2?	1	1	83.653873	<.0001*	
Location = Cluster 1?	1	1	55.029279	<.0001*	
Location = Cluster 3?	1	1	52.033652	<.0001*	
Summer?	1	1	40.793724	<.0001*	
Location = Cluster 7?	1	1	33.393565	<.0001*	
Winter?	1	1	23.356399	<.0001*	
Weekday?	1	1	16.001016	<.0001*	
Location = Cluster 6?	1	1	10.966279	0.0009*	
Light conditions = no lighting?	1	1	10.919471	0.0010*	
Casualty = baby?	1	1	7.9904765	0.0047*	
Age of driver = 70s?	1	1	7.2694484	0.0070*	
Age of vehicle = (26, 30)?	1	1	5.2470371	0.0220*	
Vehicle manoeuvre = change line?	1	1	5.0768243	0.0242*	
Road surface condition = flood?	1	1	4.414776	0.0356*	
Vehicle type = bus?	1	1	2.76359	0.0964	
Vehicle manoeuvre = turn left?	1	1	1.3803447	0.2400	
Casualty class = driver?	1	1	1.2051256	0.2723	
High_Winds_or_Not	1	1	0.4650794	0.4953	
Vehicle manoeuvre = reverse?	1	1	0.2813196	0.5958	
Casualty = aged?	1	1	0.2664465	0.6057	
Casualty = middle-aged?	1	0	0	1.0000	Removed
Age of vehicle = (16, 20)?	1	0	0	1.0000	Removed

Remove:

- Variable 1: “Casualty=Middle Age?”
- Variable 2: “Age of Vehicle=Age 16-20?”



The Adjusted Model vs Best Model So Far

Confusion Rates: Training (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.12%	10.88%
2	78.89%	21.11%

19037 rows have been excluded.

Confusion Rates: Validation (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.31%	10.69%
2	79.00%	21.00%

14219 rows have been excluded.

Confusion Rates: Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.35%	10.65%
2	78.90%	21.10%

14184 rows have been excluded.

SCREENSHOT:Best Model So Far -10.65%

Confusion Rates: Training (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.15%	10.85%
2	79.23%	20.77%

19037 rows have been excluded.

Confusion Rates: Validation (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.32%	10.68%
2	79.26%	20.74%

14219 rows have been excluded.

Confusion Rates: Test (0.1)

	Predicted	
	1	2
Accident_Severity	Row %	Row %
1	89.33%	10.67%
2	79.26%	20.74%

14184 rows have been excluded.

SCREENSHOT:Adjust Model based on Lasso -10.67%

Cross-Validation (5-fold)

Misclassification Rate Under Cross-Validation				
Fold	After LASSO (Training)	Prior LASSO (Training)	After LASSO (Validation)	Prior LASSO (Validation)
1	11.73%	11.72%	11.66%	11.64%
2	11.85%	11.46%	11.82%	11.25%
3	11.50%	11.42%	11.27%	11.41%
4	11.43%	11.72%	11.40%	11.80%
5	11.66%	11.71%	11.40%	11.78%
Average	11.63%	11.61%	11.58%	11.51%

Profit and Cost Matrix Trade-off

Bootstrap Forest for Accident_Severity

Specifications

Target Column:	Accident_Severity	Training Rows:	400454
Validation Column:	Validation 6	Validation Rows:	300341
		Test Rows:	300340
Number of Trees in the Forest:	100	Number of Terms:	34
Number of Terms Sampled per Split:	8	Bootstrap Samples:	400454
		Minimum Splits per Tree:	10
		Minimum Size Split:	1048

Overall Statistics

Measure	Training	Validation	Test	Definition
Entropy RSquare	0.0478	0.0467	0.0461	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0690	0.0674	0.0665	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3951	0.3955	0.3958	∑ -Log(ρ[j])/n
RMSE	0.3451	0.3454	0.3454	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2402	0.2401	0.2404	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1455	0.1455	0.1455	∑ (ρ[j]≠ρMax)/n
N	400454	300341	300340	n

Confusion Matrix

Training

Actual Accident_Severity	Predicted Count 1	Predicted Count 2
1	0	58280
2	0	342174

Validation

Actual Accident_Severity	Predicted Count 1	Predicted Count 2
1	0	43710
2	0	256631

Test

Actual Accident_Severity	Predicted Count 1	Predicted Count 2
1	0	43710
2	0	256630

Low Business Value

Confusion Rates: Test (0.2)

	Predicted 1	Predicted 2
Accident_Severity	Row %	Row %
1	28.23%	71.77%
2	13.69%	86.31%

Confusion Rates: Test (0.1)

	Predicted 1	Predicted 2
Accident_Severity	Row %	Row %
1	89.35%	10.65%
2	78.90%	21.10%

Main Findings

	Severity
-0.42521	Number of vehicle
0.297613	Number of casualties
0.289064	Road surface condition=dry
-0.21821	Road type=slip road
-0.21399	Road type=roundabout
0.199251	Road surface condition=wet
-0.19751	Urban
0.188765	Location Cluster 8
0.169584	Location Cluster 7
0.168999	Age of vehicle = (26, 30)?
0.110975	Location Cluster 2
0.106712	Daytime
0.090597	Road type=single carriageway
0.089613	Road surface condition=flood

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

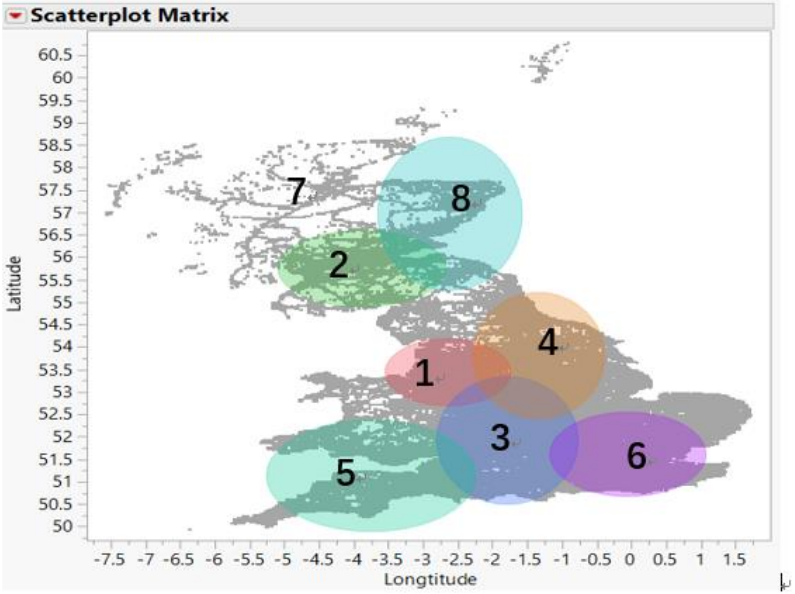
$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

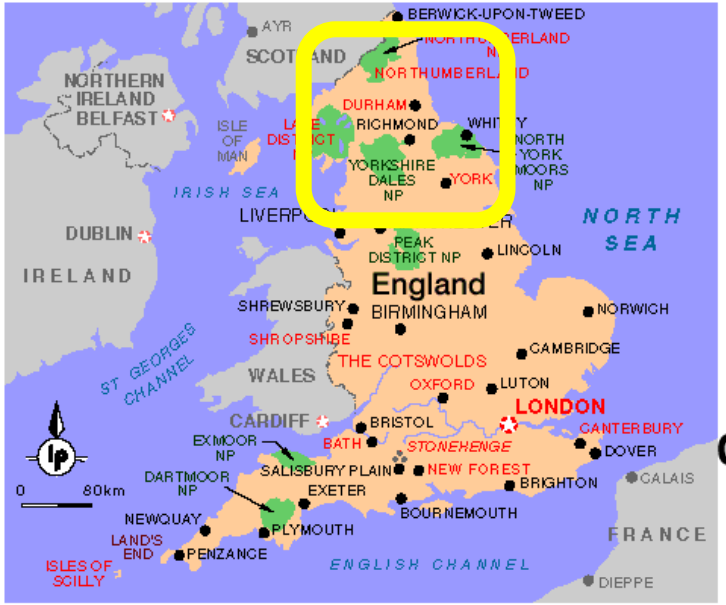
Main Findings

Variables that cause higher serious severity of accident:

Variable	Parameter
Number of casualties	0.297613
Road surface condition = dry	0.289064
Road surface condition = wet	0.199251
Location Cluster 8	0.188765
Location Cluster 7	0.169584
Age of vehicle = (26,30)?	0.168999
Location Cluster 2	0.110975
Daytime	0.106712
Road type = single carriageway	0.090597
Road surface condition = flood	0.089613

Main Findings





England

Main Findings

Variables that decrease serious severity of accident:

Variable	Parameter
Number of vehicle	-0.42521
Road type = slip road	-0.21821
Road type = roundabout	-0.21399
Urban	-0.19751

Recommendation

- Police Force Arrangement
- Insurance Company
- Drive Safety Education

Recommendation-Police Force Arrangement

- Northside areas (+)
- Aged cars (26-30) (+)
- Urban areas (-)
- Slip road and Roundabout (-)

Recommendation- Insurance Company

- Northside areas (eg. collision, medical payments, liability and uninsured coverage) (+)
- Areas have many single carriageways, or high frequency of flood (+)
- Aged cars (+)
- People who living in urban area (-)

Recommendation-Drive Safety Education

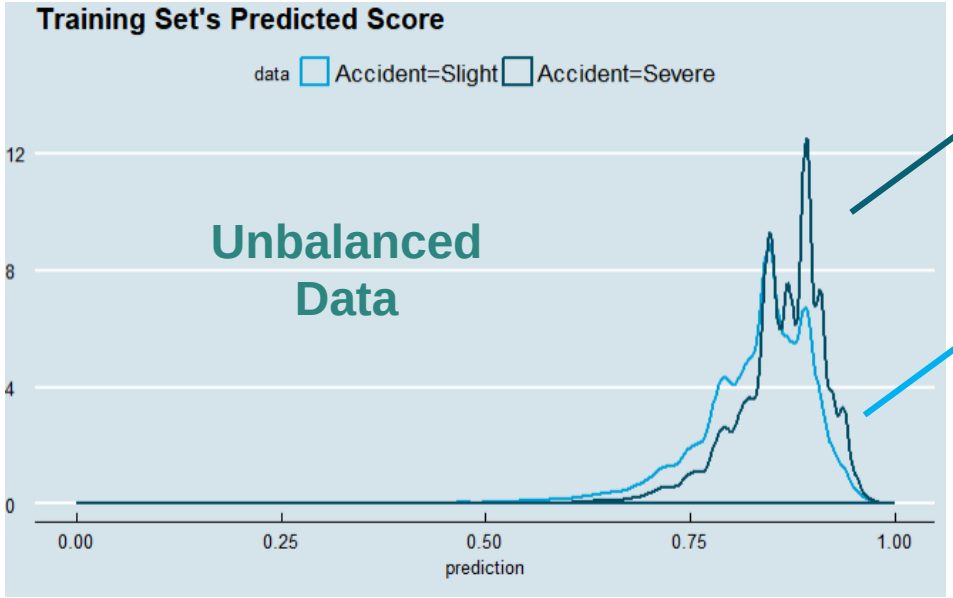
- Increase the difficulty level of driver license test in North areas
- Set more traffic signs on single carriageway
- Drive carefully on different conditions of road surface
- Replace car after the usage of 25 years

Q&A

Thank you!

Appendix: Profit and Cost Matrix Trade-off

High Scores Assigned To Sever
Low Scores Assigned To Slight



Highly right skewed

Almost the same score as “Severe”, Supposed to be on left