SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

Advanced Analytics Approach for Recall in Automobile Industry

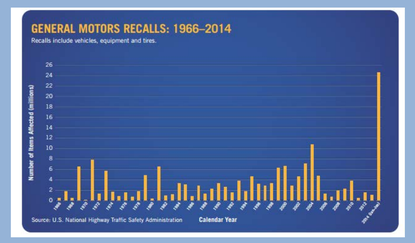
USERS PROGRAM

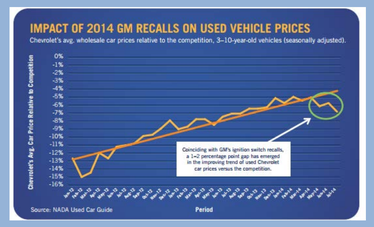
Advanced Analytics approach for recall in Automobile Industry

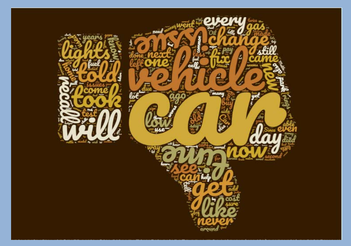
Prathap Maniyur, Prashant Nayak, Mansi Bhat

ABSTRACT

- GM had one of the biggest recalls in 2014 which involved cars such as Chevrolet Malibu, Chevrolet Impala, Monte Carlo, Cadillac CTS and Cadillac SRX[3]. In May 2014, GM had recalled more cars and trucks in 2014 than it sold in the last five years since it filed for bankruptcy. This paper is used to predict recall using social media analytics

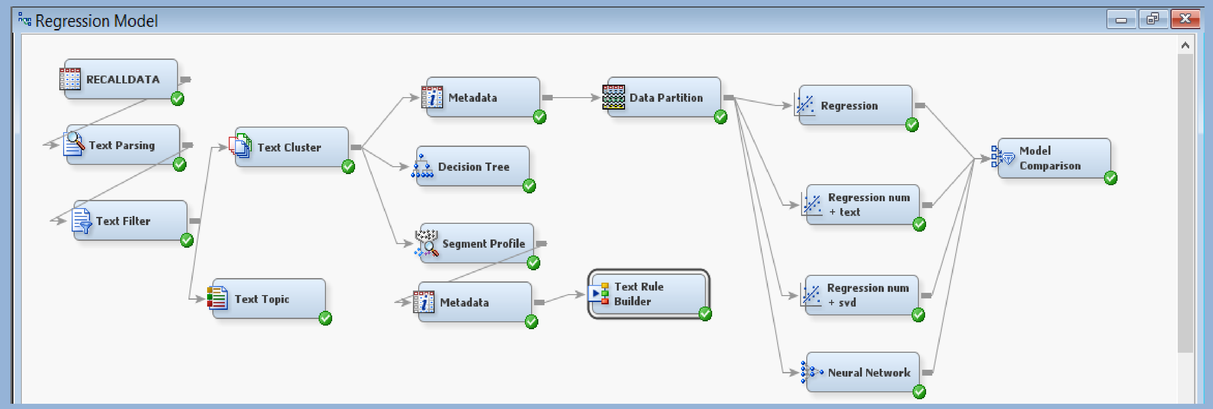






METHODS

The websites and NHTSA complaints were scrapped to collect the complaints. The unstructured data such as customer comments or feedback can enhance the power of existing predictive models. The numeric data such as mileage of the cars where the issue is reported, severity issue rating and the cost incurred for the repair is captured in the numeric and the comments are captured as text. Wanted to analyse the predictive power of the text.



RESULTS

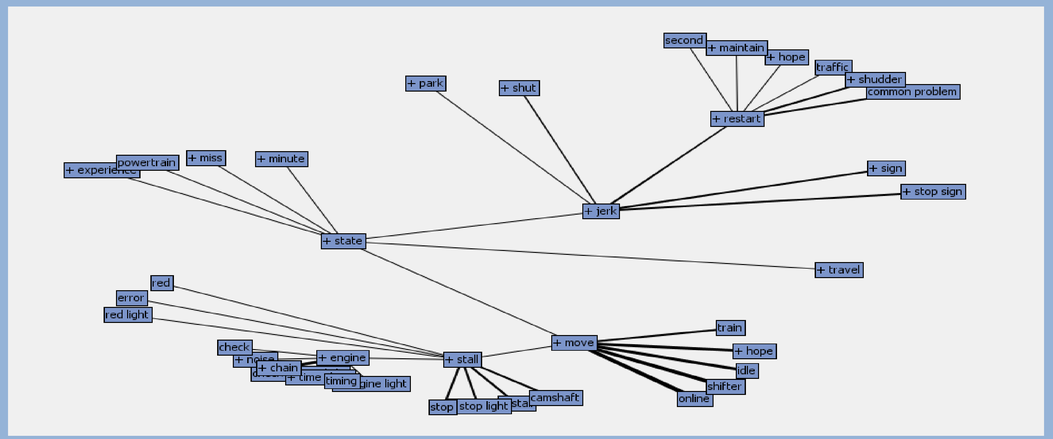
Cluster analysis

Clusters

Cluster ID	Descriptive Terms		Frequency	Percentage	Coordinate 1	Coordinate 2	Coordinate 3	Coordinate 4	Coordinate 5	Coordinate 6	Coordinate 7	Coordinate 8
	1+stall +car +happen back +recall +start chevy +drive +day dealership	...	151	67%	0.65983	-0.15211	0.145051	-0.1006	0.066538	-0.09216	-0.0112	-0.03226
	2+headlight night sierra usa terrible +truck +side gmc +road +light	...	25	11%	0.224082	0.050501	-0.19071	-0.83074	-0.10537	0.251011	-0.27004	-0.05405
	3 oil +quart +leak +mile consumption +add +check 'oil change' +buy low	...	51	22%	0.595979	0.493896	-0.04795	-0.06387	0.23995	0.10834	0.403463	-0.03617

Text topic analysis

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	11	0.150	0.059	+low,+warranty,+power,+die,home	54	14
Multiple	15	0.147	0.059	fuel,daughter,+intersection,traffic,+happen	60	19
Multiple	17	0.165	0.059	+stall,+restart,+cause,+power,+stall	51	16
Multiple	20	0.156	0.059	+shut,gmc,+steer,+back,+month	48	22
Multiple	21	0.150	0.059	+back,+park,+reverse,+hour,park	41	17
Multiple	23	0.155	0.059	+noise,+hour,+late,dept,+leave	49	30



Advanced Analytics approach for recall in Automobile Industry

Prathap Maniyur, Prashant Nayak, Mansi Bhat

RESULTS CONTINUED (CLICK TO EDIT)

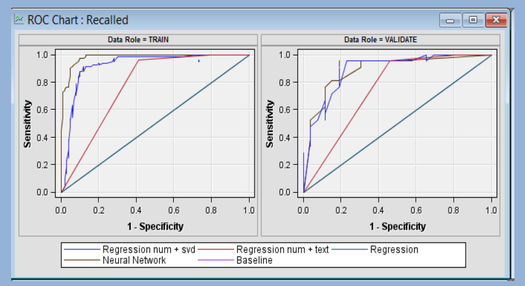
Text rule builder

Rules Obtained

Output

Target Value	Rule #	Rule	Precision	Recall	F1 score	True Positive/Total
1		1 push	100.0%	8.82%	16.22%	9/9
1		2 replace & ~usa & month	100.0%	22.55%	36.80%	15/15
1		3 cable	100.0%	27.45%	43.08%	6/6
1		4 daughter	97.22%	34.31%	50.72%	8/9
1		5 car & ~usa & ~oil	86.81%	77.45%	81.87%	64/81
1		6 stall	84.62%	86.27%	85.44%	30/42
0		7 usa	100.0%	25.60%	40.76%	32/32
0		8 quart & ~replace	100.0%	41.60%	58.76%	20/20
0		9 oil & ~stall & ~leave & ~replace	98.65%	58.40%	73.37%	38/40
0		10 night	96.51%	66.40%	78.67%	16/18
0		11 night	96.67%	69.60%	80.93%	4/4

ROC analysis



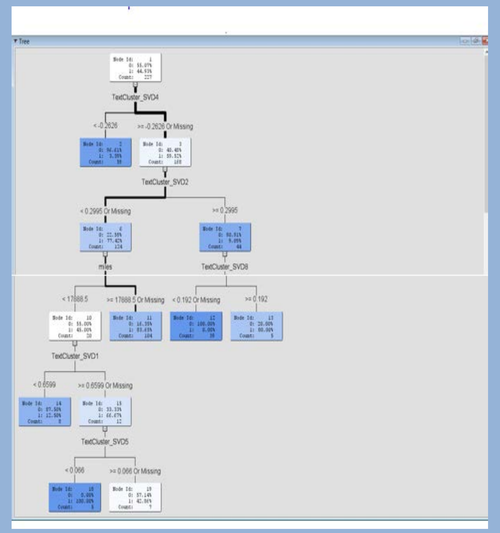
Regression Model Results

Event Classification Table
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Model Node	Model Description	Data Role	Target	Target Label	False Negative	True Negative	False Positive	True Positive
Reg3	Regression num + svd	TRAIN	Recalled	Recalled	7	87	12	74
Reg3	Regression num + svd	VALIDATE	Recalled	Recalled	2	21	5	19
Reg2	Regression num + text	TRAIN	Recalled	Recalled	3	58	41	78
Reg2	Regression num + text	VALIDATE	Recalled	Recalled	1	14	12	20
Reg	Regression	TRAIN	Recalled	Recalled	81	99	0	0
Reg	Regression	VALIDATE	Recalled	Recalled	21	26	0	0
Neural	Neural Network	TRAIN	Recalled	Recalled	5	91	8	76
Neural	Neural Network	VALIDATE	Recalled	Recalled	4	22	4	17

CONCLUSIONS

The SVDs which are statistically significant are then sent to text builder to understand the key factors, decision tree is used to get the most important factors, text cluster and text topic which in turn can provide a final set of classification rules for both recall and no recall predictions in the total data for the GM cars.



Rules Obtained

Target Value	Rule #	Rule	Precision	Recall	F1 score	True Positive/Total
1		1 car & ~oil	100.0%	60.29%	75.21%	82/83
1		2 start & ~oil	100.0%	68.21%	81.10%	47/47
1		3 hour	100.0%	74.83%	85.61%	17/17
1		4 stop	100.0%	78.81%	88.15%	17/17
1		5 month	100.0%	83.44%	90.97%	34/36
1		6 cold	100.0%	85.43%	92.14%	6/6
1		7 turn	100.0%	87.42%	93.29%	11/11
1		8 occur	100.0%	88.74%	94.04%	6/6
2		9 headlight	100.0%	64.00%	78.05%	16/16
2		10 usa & ~car & ~start	100.0%	80.00%	88.89%	15/15
3		11 quart	89.66%	50.98%	65.00%	26/29
3		12 piston	91.89%	66.67%	77.27%	10/12
3		13 leak	92.86%	76.47%	83.87%	9/8
3		14 oil & oil	93.33%	82.35%	87.50%	9/10
3		15 oil & add	91.84%	88.24%	90.00%	13/16
3		16 consume	92.00%	90.20%	91.09%	5/5

REFERENCES

1. National Highway Traffic Safety Administration(NHTSA) - Established by the Highway Safety Act of 1970 to achieve highest standards of excellence in motor vehicle and highway safety.
2. carcomplaints.com - It was started in year 2000 and was active in social media to get the buzz
3. kbb.com was launched in 1995 and is pioneer in automobile sector

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL