LEAD SCORE CASE STUDY

24.01.2022

PROBLEM STATEMENT

- To improve the conversion rate of probable customers from around 30% to above 80%.
- To identify the variables attached to the customers which are impacting the conversion rate.

MODELLING APPROACH

- Importing the dataset.
- Understanding the data
- Preparing the data for modelling
- Selecting the logistic model for the target variable
- Splitting the data to train and test dataset
- Training the model on train dataset
- Plotting the ROC curve to check the fit
- Finding optimal cut-off point for converting conversion probability to converted.
- Testing the fit with test dataset

ASSUMPTIONS MADE

- Select cell values are assumed to be Null
- Null values in Specialization variable is taken as ‘Other_Specialization’
- Null values in Tag variables is taken as Not_Tagged
- Null values in TotolVisits is taken as 1 (the least visit value)
- Null values in Page Views Per Visit is taken as 1 (the least views per visit value)
- Values with low frequencies are grouped together and assumed to be same significance

FINAL MODEL OUTPUT

	coef	std err	z	P>|z|	[0.025	0.975]
const	-3.8106	0.226	-16.841	0.000	-4.254	-3.367
Welingak Website	5.5386	1.019	5.438	0.000	3.542	7.535
Modified	-1.1340	0.122	-9.276	0.000	-1.374	-0.894
SMS Sent	2.2480	0.124	18.119	0.000	2.005	2.491
Busy	3.4552	0.305	11.338	0.000	2.858	4.052
Closed by Horizzon	9.3948	0.747	12.572	0.000	7.930	10.859
Lost to EINS	8.8270	0.751	11.759	0.000	7.356	10.298
Not_Tagged	2.4663	0.226	10.909	0.000	2.023	2.909
Other_Tag	3.3572	0.546	6.149	0.000	2.287	4.427
Ringing	-0.9309	0.311	-2.993	0.003	-1.540	-0.321
Will revert after reading the email	7.0587	0.268	26.345	0.000	6.534	7.584
in touch with EINS	3.3013	0.819	4.031	0.000	1.696	4.906
switched off	-1.2865	0.560	-2.296	0.022	-2.385	-0.188

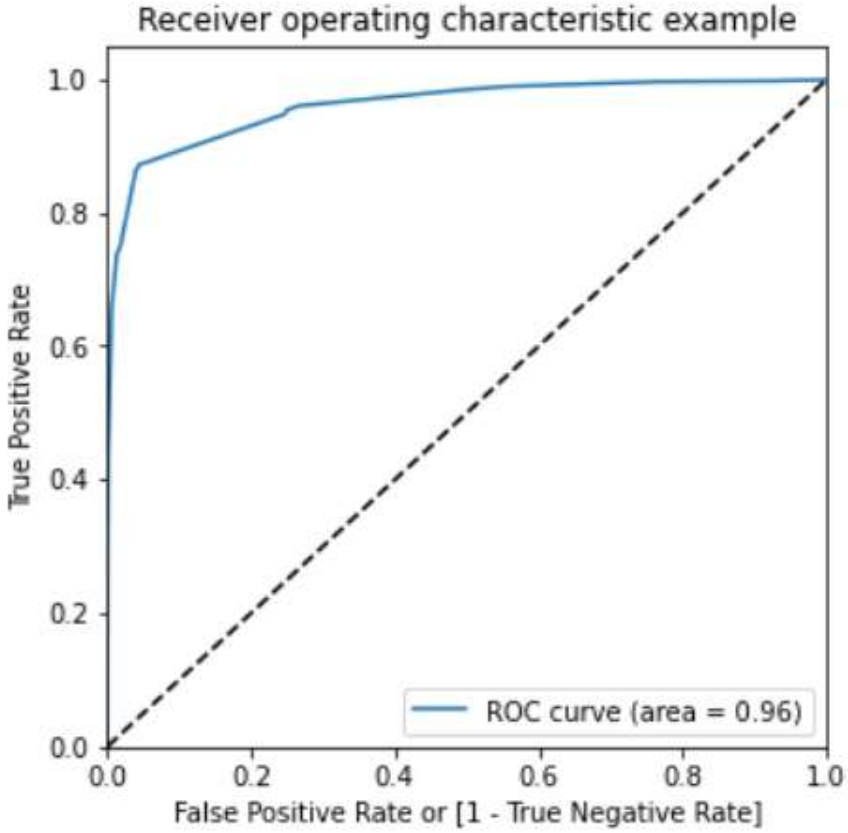
All P values are below .05 as required

VIF(S) OF THE FINAL MODEL

	Features	VIF
4	Closed by Horizzon	1.08
0	Welingak Website	1.04
3	Busy	1.04
5	Lost to EINS	1.04
11	switched off	1.04
7	Other_Tag	1.00
10	in touch with EINS	1.00
2	SMS Sent	0.26
9	Will revert after reading the email	0.19
8	Ringing	0.05
1	Modified	0.03
6	Not_Tagged	0.03

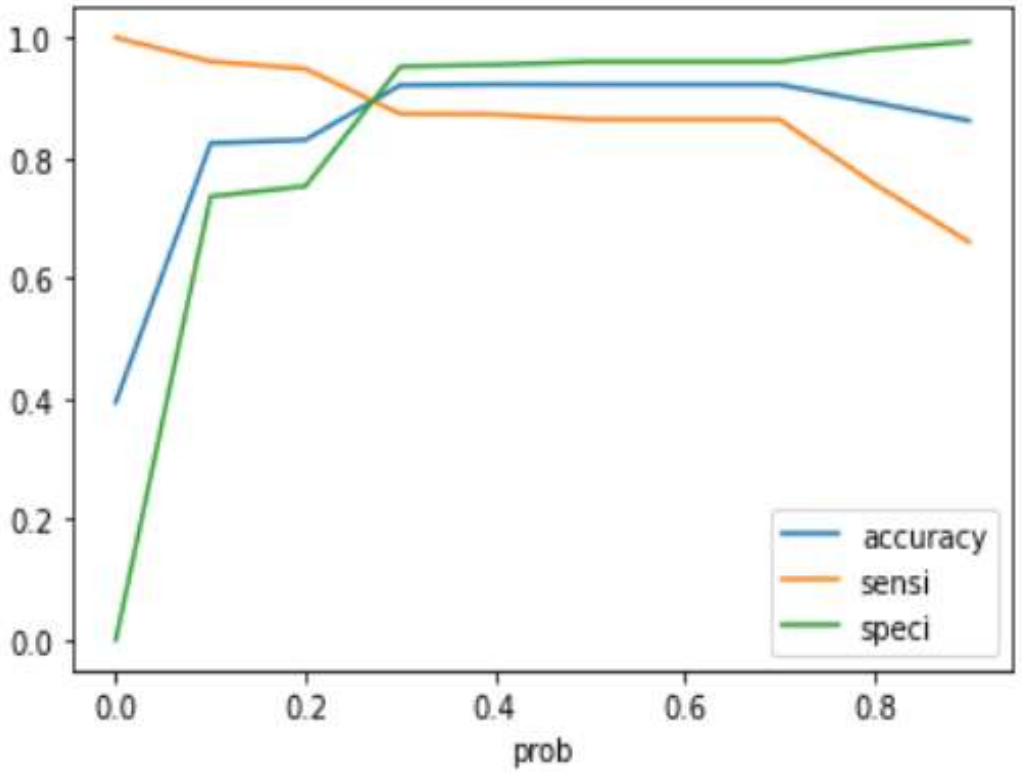
All VIF values are below 5 as required

ROC CURVE
(TRADE-OFF BETWEEN SPECIFICITY AND SENSITIVITY)



The ROC curve is straight up and close to the True Positive Rate axis as required

SELECTING THE OPTIMAL CUT-OFF POINT FOR PROBABILITY CONVERSION



The cut-off point can be seen at the probability level of .28

ANALYSIS

- Accuracy Score of Test Model is .93
- Sensitivity of Test Model is .88
- Specificity of Test Model is .95

The above values are close to the train data set values which are .92, .87, .95 respectively.

CONCLUSION AND RECOMMENDATIONS

- Based on the model output it is identified that ‘Closed by Horisson’, ‘Lost to EINS’ and Will revert after reading the email are the top three variables effecting the conversion rate.
- Based on the original Variable names, the variable Tags, Lead Source and Last Notable Activity are most important.
- It is recommended that the company should focus on customers who are tagged Closed to Horizzon, Lost to EINS or will revert after reading the email along with last Notable Acitivity as SMS sent and Lead Source as Welingak website.

THANK YOU