

1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Answer:

The top three variables that contribute the most are

- \* TotalVisits
- \* Tags
- \* Lead Origin

Reviewing coefficient of these variables indicate the importance of these variables.

Generalized Linear Model Regression Results						
=====						
Dep. Variable:	Converted	No. Observations:	4473			
Model:	GLM	Df Residuals:	4460			
Model Family:	Binomial	Df Model:	12			
Link Function:	Logit	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-1050.9			
Date:	Sun, 01 Jan 2023	Deviance:	2101.7			
Time:	17:15:52	Pearson chi2:	1.06e+04			
No. Iterations:	7	Pseudo R-squ. (CS):	0.5990			
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----	-----	-----	-----	-----	-----	-----
const	-2.8431	0.151	-18.781	0.000	-3.140	-2.546
TotalVisits	15.2669	3.860	3.955	0.000	7.702	22.832
Total Time Spent on Website	3.7198	0.253	14.694	0.000	3.224	4.216
Lead Origin_Lead Add Form	4.5655	0.281	16.251	0.000	4.015	5.116
Do Not Email_Yes	-1.8817	0.261	-7.209	0.000	-2.393	-1.370
Last Activity_SMS Sent	1.3417	0.143	9.411	0.000	1.062	1.621
Last Notable Activity_Modified	-0.6904	0.125	-5.513	0.000	-0.936	-0.445
Last Notable Activity_Olark Chat Conversation	-2.0413	0.470	-4.345	0.000	-2.962	-1.120
Tags_Busy	1.3778	0.240	5.750	0.000	0.908	1.847
Tags_Lost to EINS	5.9163	0.748	7.907	0.000	4.450	7.383
Tags_Ringing	-2.9789	0.270	-11.035	0.000	-3.508	-2.450
Tags_Will revert after reading the email	3.5337	0.134	26.396	0.000	3.271	3.796
Tags_switched off	-3.3143	0.734	-4.517	0.000	-4.752	-1.876
=====						

2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

Answer:

The top 3 categorical/dummy variables in the model that should be focused the most on in order to increase the probability of lead conversion are

- Tags – Status “switched off “
- Tags – Status “Ringing”
- Last Notable Activity - “Olark Chat Conversation”

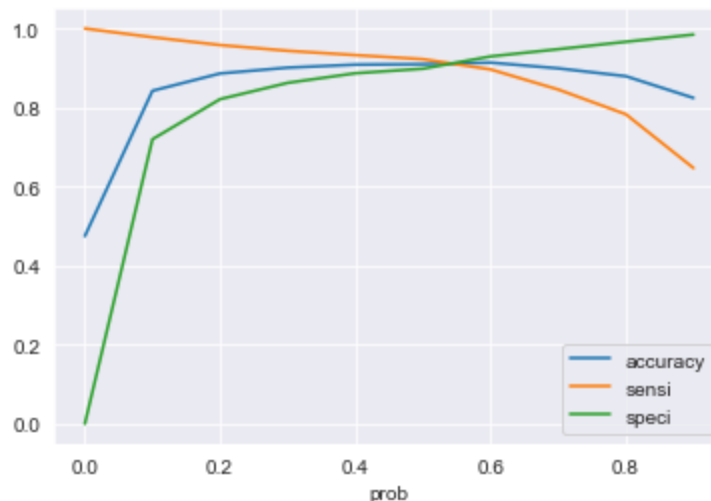
3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

Answer:

In order to make the sales aggressive, the company must also contact the leads which have a conversion probability (value = 1) under the cut off 0.55.

Here, the concept of sensitivity is required. ***Sensitivity = True Positives / (True Positives + False Negatives)*** With respect to our model, sensitivity can be defined as the number of actual conversions predicted correctly out of total number of actual conversions.

Different values of sensitivity can be achieved for the model by changing the cutoff threshold for probability of lead conversion. For our model, below is the graph showing changes in Sensitivity, Specificity and Accuracy with change in the threshold:



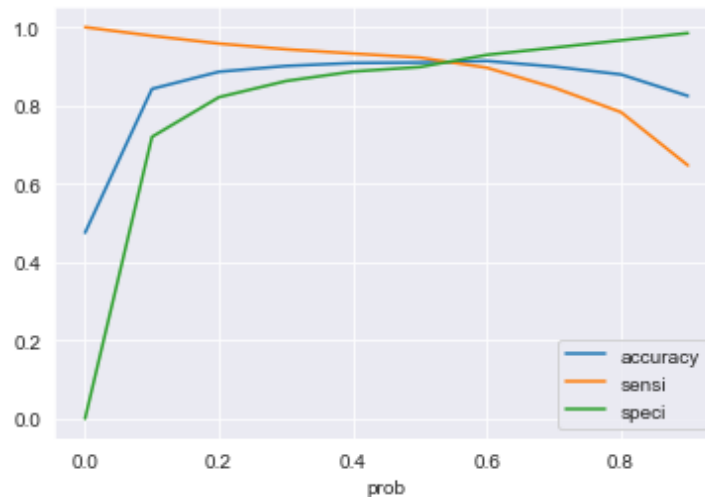
As we can see, sensitivity decreases as the threshold increases. In the given situation, we'll need a high sensitivity because high sensitivity will mean that our model will correctly predict almost all leads who are likely to convert. At the same time, it may overestimate and misclassify some of the non-conversions as conversions. But as the company has extra **man-power** for two months and wants to make the lead conversion more aggressive by making phone calls to as much potential leads as possible, it is a good strategy **to go for high sensitivity**. To achieve high sensitivity, we need to choose a **low threshold value**.

4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

Answer:

In order to minimize the rate of useless phone calls, the company may contact all the leads which have a conversion probability (value = 1 highlighted in yellow color) under column 0.7.

Here, the concept of specificity is required. ***Specificity = True Negatives / (True Negatives + False Positives)*** With respect to our model, specificity can be defined as the number of actual non-conversions predicted correctly out of total number of actual non-conversions.



From the above graph, we can see that the specificity increases as the threshold increases. In the given situation, we'll need a **high specificity** because high specificity will mean that our model will correctly predict almost all leads who are not likely to convert.

At the same time, it may misclassify some of the conversions as non-conversions. But as the company has already reached its target for a quarter and doesn't want to make phone calls unless it's extremely necessary, it is a good strategy to go for high specificity.

It will ensure that the phone calls are only made to customers who have a **very high probability of conversion**. To achieve high specificity, we need to **choose a high threshold value**.