

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

The modelling summary is given below.

	coef	std err	z	P> z	[0.025	0.975]
const	-1.5576	0.102	-15.201	0.000	-1.758	-1.357
Total Time Spent on Website	3.9530	0.134	29.560	0.000	3.691	4.215
Lead Origin_Lead Add Form	3.3976	0.171	19.864	0.000	3.062	3.733
Lead Source_Direct Traffic	-0.6349	0.073	-8.694	0.000	-0.778	-0.492
Lead Source_Welingak Website	2.5704	1.022	2.514	0.012	0.567	4.574
Do Not Email_Yes	-1.2497	0.151	-8.295	0.000	-1.545	-0.954
Last Activity_Email Opened	0.2470	0.094	2.634	0.008	0.063	0.431
Last Activity_Had a Phone Conversation	2.1333	0.547	3.902	0.000	1.062	3.205
Last Activity_SMS Sent	1.4542	0.093	15.679	0.000	1.272	1.636
Specialization_Unknown	-0.3976	0.073	-5.441	0.000	-0.541	-0.254
Last Notable Activity_Modified	-0.8791	0.078	-11.248	0.000	-1.032	-0.726
Last Notable Activity_Olark Chat Conversation	-0.7842	0.286	-2.745	0.006	-1.344	-0.224
Last Notable Activity_Unreachable	1.6262	0.462	3.520	0.000	0.721	2.532

The features with highest absolute value are the most important. In the above report, the following 3 features have the highest absolute coefficients

- a. Total time spent on Website
- b. Lead origin(Lead add form)
- c. Lead source(Welignak website)

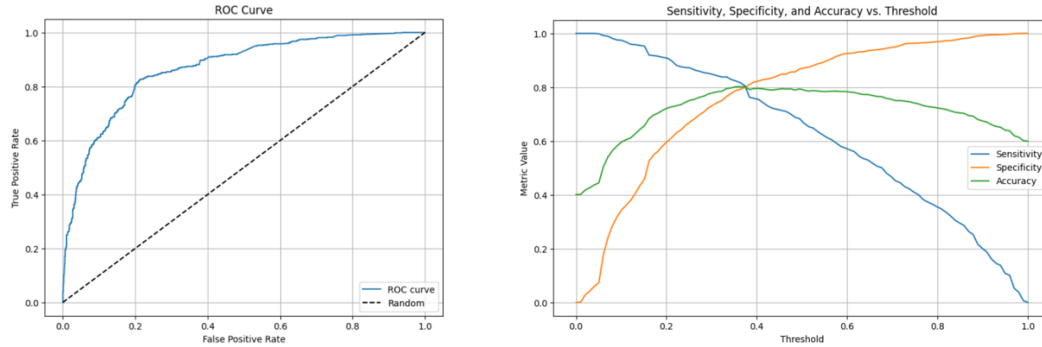
Hence, leads that are spending more time on the website, filling forms on Landing page submission or coming from Welingak website should be targeted for improving conversion.

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?
 - a. Lead Origin(Lead add form)
 - b. Lead Source_Welingak Website
 - c. Last Activity_Had a Phone Conversation

Thus, Leads filling up the form on landing page or coming from Welingak website or just had a phone call conversation will be most likely to convert

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.

The model's ROC curve and sensitivity, accuracy and specificity curve is given below:



As per the curve above, the optimal probability point that gives 82% sensitivity and balances against other metrics will be 0.35. The model output gives probability of conversion along with the prediction of converted or not.

The model can be employed in the following ways. The increased workforce can help in acquiring more customers. To increase the top of the funnel, the chosen threshold of 0.35 can be lowered. If the workforce increases by X%, the threshold should be chosen in such a way that the number of leads for the sales team also increases by X%. This would have an impact on accuracy for the 2 months but will lead to more conversions to make use of increased workforce and the model output still helps us to quantify the conversion rate of a new lead. Hence, they should focus on the high scoring leads from the logistic regression model.

Increased workforce can also have more prompt follow-ups using SMS, chat, and email.

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. –

During this period, X education can tune the model to only provide very high scoring leads(predicted as converted==1 by the model and having very high probability output). The calls can be skipped in favour of other mediums like chat, email and SMS. Using this, they will only be calling after the customers have bought the course or are facing problem in buying the course.