

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

The top 3 variables were that increase the probability of conversion are:

1. Total time spent on website (numeric variable)
2. Page views per visit (numeric variable - negatively related)
3. Lead Quality (categorical variable)

	coef	std err	z	P> z	[0.025	0.975]
const	-1.1036	0.197	-5.592	0.000	-1.490	-0.717
Lead Origin	0.5277	0.075	7.007	0.000	0.380	0.675
Lead Source	0.1451	0.014	10.032	0.000	0.117	0.173
Do Not Email	-1.4634	0.177	-8.263	0.000	-1.810	-1.116
Total Time Spent on Website	4.2489	0.169	25.100	0.000	3.917	4.581
Page Views Per Visit	-4.7586	0.505	-9.414	0.000	-5.749	-3.768
Last Activity	0.1253	0.010	12.001	0.000	0.105	0.146
What matters most to you in choosing a course	-0.2742	0.032	-8.447	0.000	-0.338	-0.211
Tags	0.0759	0.006	12.047	0.000	0.064	0.088
Lead Quality	-0.9228	0.036	-25.393	0.000	-0.994	-0.852
City	0.0643	0.020	3.152	0.002	0.024	0.104
A free copy of Mastering The Interview	-0.2953	0.094	-3.139	0.002	-0.480	-0.111

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

The top 3 categorical variables were that increase the probability of conversion are:

1. Do not email
2. Lead Quality
3. Lead Origin

4. 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.

In terms of the model, we need to increase 'Recall.

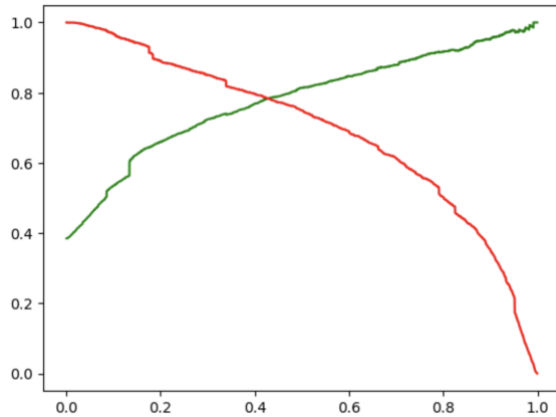
Precision means of those detected as converting lead, how many converted. In other words, it is the measure of the model's ability to correctly predict the conversion leads.

Recall on the other hand means of those leads which converted, how many did the model predict.

The graph below shows a precision-recall tradeoff curve.

The red line denotes the recall, that decreases as the cut-off threshold increase.

The green line denotes precision, that increase as the cut-off threshold decreases.



5. 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.

In terms of the model, we need to increase 'Precision'.

In the current situation where the company focus is to capture as many potential conversion leads as possible, we prioritize recall.

In the case where the company wants to reduce the number of phone calls it wants to make, then accurately detecting positive leads would be a priority. Thus high precision, high cut-off threshold (between 0.6 – 0.8).