

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

Ans. The top three variables that contribute positively towards the leads getting converted are: -

- a) Lost to EINS (Tags)
- b) Total Time Spent on website
- c) Lead Add Form (Lead Origin)

The top three variables that contribute positively towards the leads getting converted are: -

- a) Already a Student (Tags)
- b) Ringing (Last Activity)
- c) Null Count (Number of missing values for a lead)

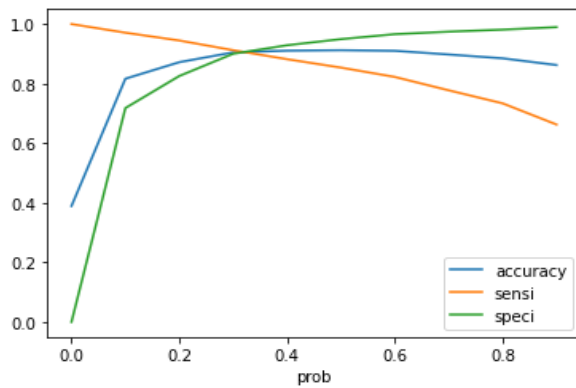
2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

Ans. The top three dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are: -

- a) Lost to EINS (Tags)
- b) Lead Add Form (Lead Origin)
- c) Will Revert After Reading the Email (Last Activity).

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.

Ans. We plotted a curve for sensitivity, specificity and accuracy and decided 0.3 to be the optimal cutoff, i.e. the lead should be labeled as converted if the lead probability according to the model is more than 0.3. Even though, the sensitivity of our model is more than 90%, if X-Education wants to be even more aggressive, they could set the optimal cutoff to be 0.2 (94% sensitivity) and the interns should give call to all the leads that get the probability of conversion is more than 0.2. In addition to calls, they should send emails and contact the leads on olark chat. The cutoff graph is given below.



	prob	accuracy	sensi	speci
0.0	0.0	0.389207	1.000000	0.000000
0.1	0.1	0.816453	0.970997	0.717975
0.2	0.2	0.871965	0.944776	0.825570
0.3	0.3	0.904902	0.912197	0.900253
0.4	0.4	0.910623	0.882002	0.928861
0.5	0.5	0.912169	0.853794	0.949367
0.6	0.6	0.910159	0.822408	0.966076
0.7	0.7	0.897789	0.777513	0.974430
0.8	0.8	0.884954	0.733810	0.981266
0.9	0.9	0.862378	0.662694	0.989620

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.

Ans. If the company wants to minimize the useless phone calls, the company should focus on increasing the specificity. According to the given table, the optimal cutoff should be increase to 0.5 to increase the specificity to 0.95. Even though the dip in the sensitivity isn't large, but still, if the company wants to avoid useless calls, the company should still mail the leads as they can mail the leads in bulk without giving attention to each one of the lead. This would also result in less dip in sensitivity.