

## Lead Scoring Case Study Subjective Question and Answers

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

Based on coefficient values and VIF of different variables, top three variables that contribute most in getting lead score probability are:

- Tags\_Lost to EINS
- Tags\_Closed by Horizzon
- Tags\_Will revert after reading the email

Following are the coefficient values and VIFs of the selected features:

Feature Name	Coefficient value	VIF value
<b>Tags_Lost to EINS</b>	7.2832	1.06
<b>Tags_Closed by Horizzon</b>	6.3230	1.22
<b>Tags_Will revert after reading the email</b>	4.6731	0.06

### 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?

Like mentioned in previous solution following features are the top 3 categorical features that should be focused most in order to increase the probability of lead conversion:

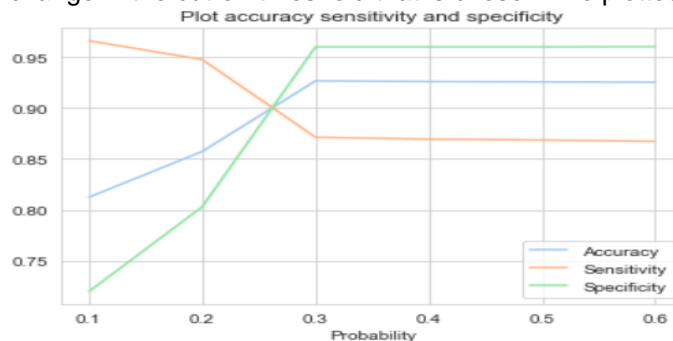
- Tags\_Lost to EINS
- Tags\_Closed by Horizzon
- Tags\_Will revert after reading the email

### 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.

We have included specificity and sensitivity in our model validation metrics, which if focused on will be a game changer in getting more conversions.

In consideration of our model, Sensitivity is the ratio of total number of actual conversions correctly predicted as total number conversions (True Positives). On the other hand Specificity can be defined as total number of actual non-conversions predicted as actual-conversions (True Negatives).

If anyone from specificity and sensitivity increases other will decrease and this happens because of change in the cut-off threshold that is chosen. We plotted one graph to describe this scenario.



As we can see when the cut-off threshold is very low sensitivity is high while specificity is very low and with increase in cut-offs threshold sensitivity is getting low while specificity is getting higher.

If our model has high sensitivity, then it will be able to correctly identify all the leads that are likely to convert and this can be achieved by choosing a lower cut-off threshold.

Now that we know X Education has greater manpower for 2 months they can reach out to all the leads predicted by our model as after decreasing the threshold which will in turn increase sensitivity then we will have a large number of leads with higher chances of conversion. This will ensure the increase in conversion rate if followed.

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.

Since the company has reached its target for the quarter and does not want to spend unnecessary time on phone calls which requires a lower number of leads. For this we can use the reverse of the approach that we used in last solution. If we increase our model's cut-off threshold then it will result in higher specificity which means it will predict almost all the leads which are not likely to convert. As a result this will give a lesser number of leads that have chances of converting, saving a lot of time and phone calls by the sales team.