LEAD SCORE ASSIGNMENT

Assignment based questions:

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

1. Based on the coefficient values the final model, the following are the top three variables that contribute most towards the probability of a lead getting converted:

a) Working professional (from what is the current occupation)

b) Other_References (from Lead Source)

c) Others Activity -SMS sent (from the Lead Notable Activity)

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?

2. The top 3 categorical variables are :

a) Working professional (from what is the current occupation)

b) Other_References (from Lead Source)

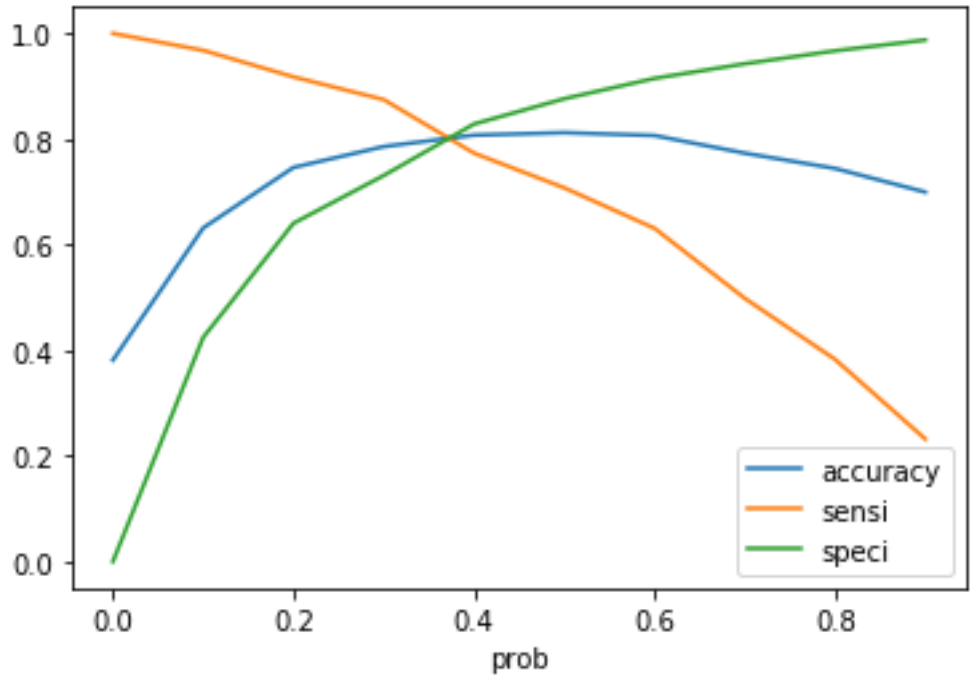
c) Others Activity -SMS sent (from the Lead Notable Activity)

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.

3.Sensitivity with respect to our model can be defined as the proportion of actual Conversions that are correctly predicted .Similarly, Specificity can be defined as the proportion of actual non-Conversions that are identified.

If we consider a specific model, with the increase in sensitivity, the specificity decreases and vica versa. Thus, the different values of the sensitivity and specificity can attain for the same model by changing the Conversion Probability cut-off threshold value.

For example, in our model when the probability thresholds are very low, the sensitivity is very high and specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high. And at about 0.4, the three metrics seem to have decent values (i.e accuracy 80.7%,sensitivity of 77%,and specificty of 82%) hence, we choose 0.4 as the optimal cut-off point.



High sensitivity means that the model will correctly identify almost all leads who are likely to Convert. Now, since X Education has more man-power for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all of the potential leads, the strategy would be to go for lower threshold value for Conversion Probability. This will ensure the Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible.

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.

4. In context of the argument explained in the previous question, we know that the high specificity means that the model will correctly identify almost all leads who are not likely to Convert. Thus, in this case where the company's aim is to not make phone calls unless it's extremely necessary, the strategy should be to have higher threshold value for Conversion Probability. This will ensure the Specificity value is very high, which means that almost all leads almost all leads who are not likely to Convert are identified correctly and as a result the agents won't have to make unnecessary phone calls and can focus on some new work