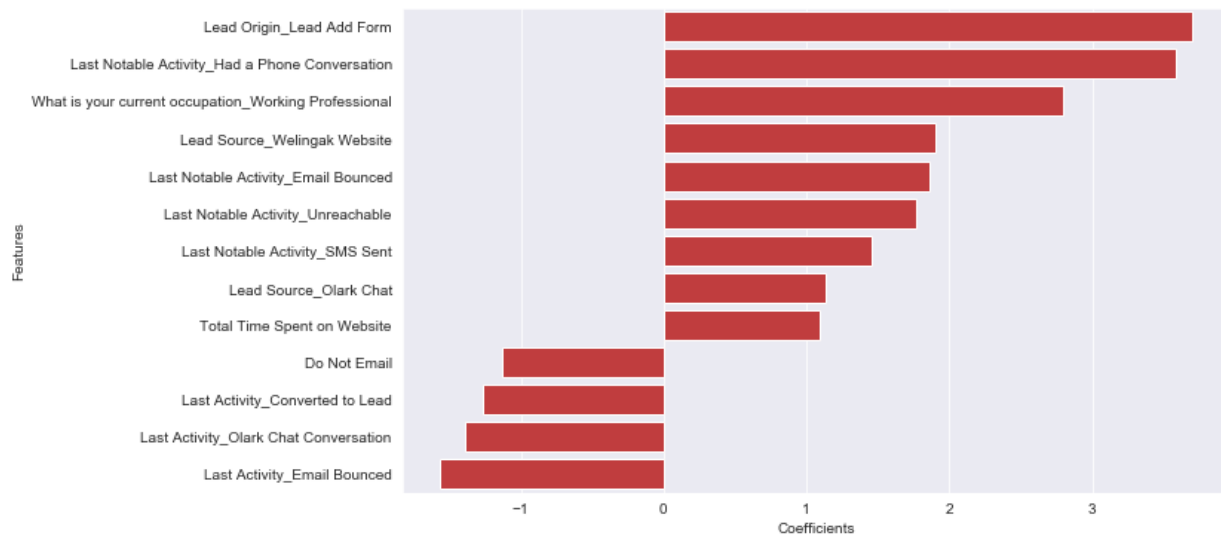


1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Ans:

The variables used to build the model is represented below in the basis of their variable coefficient which importance in the lead conversion



From the above bar graph, the top 3 variables that contribute the most towards the probability of a lead getting converted are:

- Lead Origin_Lead Add From
- Last Notable Activity_Had a Phone Conversation
- What is your current occupation_Working Professional

-
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: From the above bar graph, are the top 3 categorical/dummy variables that contributes the most in order to increase the probability of lead conversion are:

- Lead Origin_Lead Add From
 - Last Notable Activity_Had a Phone Conversation
 - What is your current occupation_Working Professional
-

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:

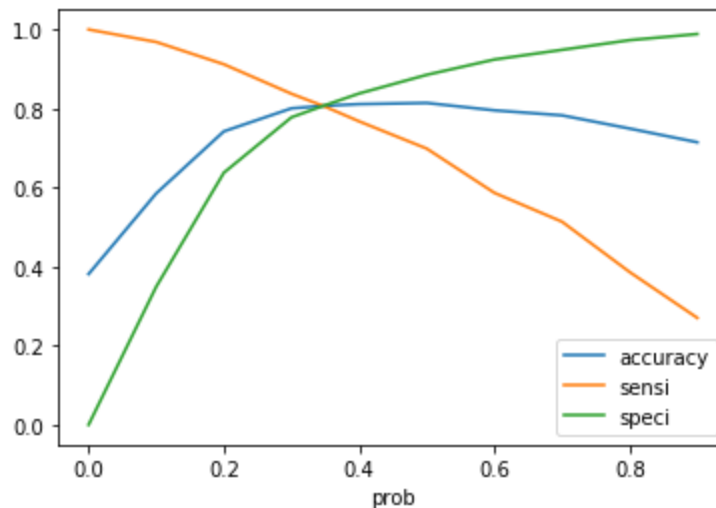
Sensitivity in our model present can be defined as the ratio of 'Number of actual Yeses correctly predicted' by 'Total number of actual Yeses'

$$\text{Sensitivity} = \frac{\text{Number of actual Yeses correctly predicted}}{\text{Total number of actual Yeses}}$$

Specificity can be defined as the ratio of 'Number of actual Nos correctly predicted' by 'Total number of actual Yeses'

$$\text{Specificity} = \frac{\text{Number of actual Nos correctly predicted}}{\text{Total number of actual Yeses}}$$

From our model, the below graph which shows the Sensitivity and Specificity which changes with the change in the threshold value:



From this graph above, we conclude that the optimal cut-ff for the model is around 0.35 and we chose this value to be our threshold value and we get a decent value of all the three matrices – Accuracy (~80.71%), Sensitivity (~80.45%) and Specificity (~80.86%). When the threshold value is very low Sensitivity is very high and specificity is very low. Similarly, for larger threshold value, the sensitivity value is very low and specificity value is very high.

High sensitivity implies that our model will correctly identify almost all the leads which are likely to get converted. This means that our model is overfitting i.e., it will misclassify some of non-conversion leads as Convert.

Now, as X Education has the manpower for the period of 2 months, and they wish to make the lead conversion more aggressive by wanting almost all the potential leads to be converted. By **choosing a lower threshold value for Conversion rate** they can achieve the required lead conversion rate.

By choosing a lower threshold value, it ensures that the Sensitivity value will be very high which in turn will make sure that almost all leads who are likely to Convert are identified correctly and the sales team of X Education can make phone calls to as much of 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.

Ans:

High sensitivity implies that our model will correctly identify almost all the leads which are likely to get converted. This means that our model is overfitting i.e., it will misclassify some of non-conversion leads as Convert.

So, since X Education has already reached its target for a quarter before the deadline and they want to minimize the rate of useless phone calls unless its extremely necessary. We can **choose a high threshold value for Conversion rate** as it implies that the Specificity value will be very high and Sensitivity value will be very low which will make sure that all the leads who were at the edge of getting converted or are not selected. By this sales teams will not have to make unnecessary phone calls and focus on some other work.
