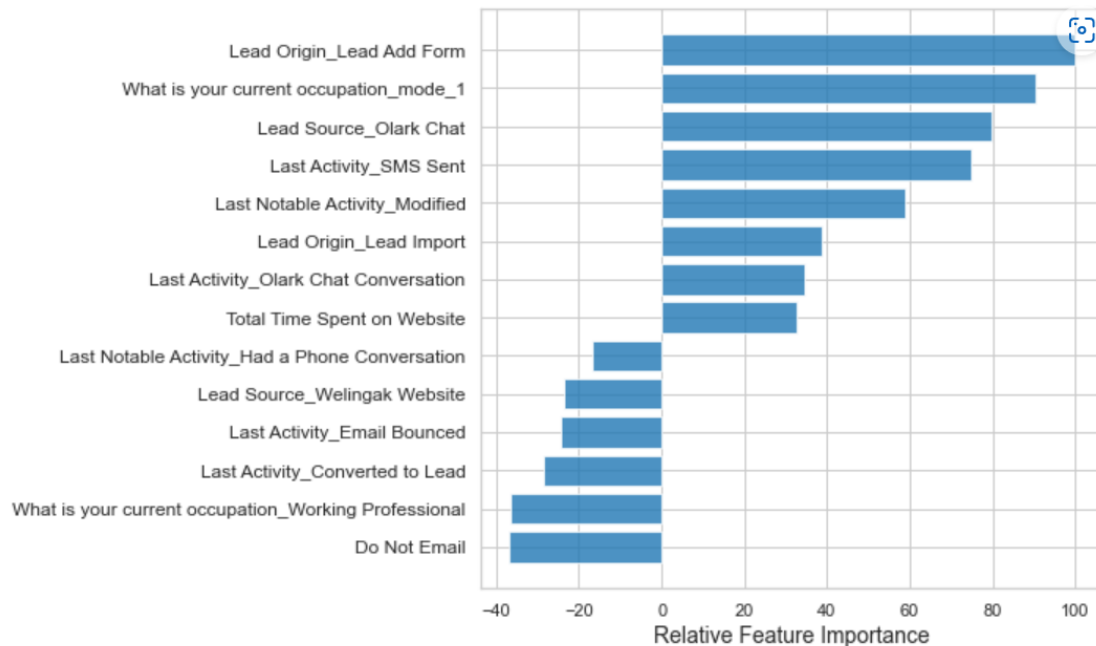1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?
   Answer :
   The features used to build the model have been represented below based on their importance in lead conversion as per their coefficient values.

   Following three features have biggest coefficient which directly correspond to whether lead is converted or not:
     1) **'Lead Origin' where value is 'Working Professional'**
     2) **'Last Notable Activity' where the value is 'Lead Add Form'**
     3) **'Lead Source' where the value is 'Welingak Website'**

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?
   Answer:
   Following dummy variables should be focussed in order to increase lead conversion:
     1. Lead Origin_Lead Add Form
     2. Last Notable Activity_Had a Phone Conversation
     3. 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.

Answer:
We would require a model with very good sensitivity so that it predicts the people who would be converted successfully accurately and contact them.

Sensitivity = True Positives/ (True Positives + False Negatives)

With respect to our model, sensitivity can be defined as the number of actual conversions predicted correctly out of total number of actual conversions. Different values for sensitivity can be achieved for the model by varying the cutoff threshold for probability of lead conversion.

For our final model, below is the graph showing varying values of Sensitivity, Specificity and Accuracy with change in the threshold.
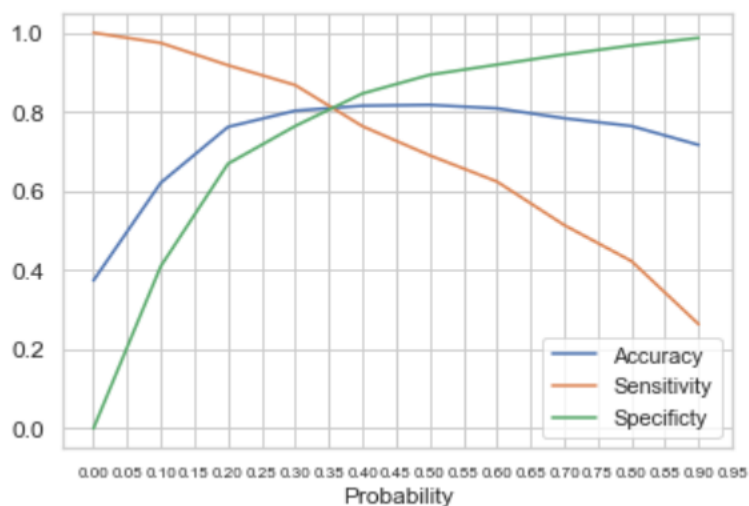
As visible in the plot, sensitivity decreases as the threshold increases. In the given situation, we'll need a high sensitivity because high sensitivity will mean that our model will correctly predict almost all leads who are likely to convert.

At the same time, it may overestimate and misclassify some of the non-conversions as conversions. But as the company has extra man-power for two months and wants to make the lead conversion more aggressive by making phone calls to as much potential leads as possible, it is a good strategy to go for high sensitivity. To achieve high sensitivity, we need to choose a low threshold value.

Additionally, a good strategy for the company would be to contact people who have 'Lead Add Form' value as 'Lead Origin'. And the students who have contacted latest via phone and had conversation. And additionally, the students should be working professional. Since candidate having these will have highest chances of being converted.

And it would be best to avoid students who have chosen 'yes' for 'do not email' option and who have not provided any input for the current occupation. It will be least likely for

them to be converted.

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.
   Answer:
   In order to focus on some new work and to avoid unnecessary calls, we will primarily have to identify the candidates who are not likely to convert. Meaning we will have to predict which candidates have least chances of conversion and we can primarily focus on ways to convince them in order to get some new work.

   In order to get this result we will have to increase our model specificity accordingly by selecting appropriate cutoff so that candidates who wont be converted are predicted accurately.

   As per the graph shown in previous answer, the specificity of our model will increase as we increase the threshold cutoff for our model. It will identify almost all leads who are not likely to be converted and can be focused upon for new work.