Q1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Top three variables in our model is if we are considering constant also as a variable are in following order:-

1. origin_Lead Add Form 4.7071
2. constant -1.6458
3. last_notable_activity_SMS sent 1.5859

Top three variables in our model is if we are not considering constant also as a variable are in following order:-

1. origin_Lead Add Form 4.7071
2. last_notable_activity_SMS sent 1.5859
3. last_activity_Email Bounced -1.5312

Q2. 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?

Origin Lead Add form is something not in our control, we should focus on what actions on organization end result in more lead conversion. As per our model -

1. Lead source Olark chat has a positive coefficient of 1.4738. So we should try to put more people on olark chat to generate the lead source from there.
2. Last activity olark chat conversation has negative coefficient of -1.3043. It means a lot of our leads drop off after the olark chat conversation. So we need to find out if our employees are not trained enough who are there on olark chat and decrease the drop off from there itself.
3. Email bounced has a negative coefficient of -1.5312. So it means we are not validating the email and entering the wrong email on the lead form. Let it be website or past referrals. On website we should check if email validation is there before form submission.

Q3. 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.

This means that they have a lot of workforce. The good strategy for the company will be not to miss any lead which can be converted whereas it is ok that some leads which wont be converted gets the phone call.

So we will have to increase
1.Sensitivity - No of actual leads predicted compared to actual no of hot leads.
2. Recall - Probablity that a actual lead is predicted correctly meaning that we dont classify hot lead in the cold lead category.

Also we will have to balance Specificity so that not all the people gets call.

In our model at cutoff = 0.2, we have good values. So we should look at this cutoff point during this period.

sensitivity = 0.8971373119844736

specificity = 0.633361087982237

Q4. 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.

At this time we want to focus. We have to make less phone calls but all the leads should get converted. So we will have to focus more on specificity and precision(probablity that a predicted lead is actual lead). We can compromise on sensitivity of the model.

In our model at cutoff = 0.6 So we should look at this cutoff point during this period.
sensitivity = 0.6016496846191169
specificity = 0.9214543436025534
precision = 0.814182534471438

So almost 60% of our hot leads we are able to capture. 90% of the total non hot leads we are able to capture. Also 81% time a predicted hot lead is actual a hot lead.