

Answers to the subjective questions

1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?
 - The top three variables based on coefficients and also from exploratory data analysis are:
 - a. Lead Source
 - b. Total Time Spent on Website
 - c. What is your current occupation
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?
 - The top three categorical/dummy variables based on coefficients are:
 - a. Lead Source_Welingak Website
 - b. Lead Source_Reference
 - c. 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.
4. We need to want to make phone calls to as much of such people as possible by the interns, so, we can have high false positives. To address, we can lower the cut-off on the prediction probability. We could observe that lowering the cutoff from 0.42 to 0.2 increases the number of false positives from 206 to 533 which has more than doubled. So, we can decrease the cut-off to 0.2 and get the interns calling all the contacts with positives.
5. 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.
 - As we increase the cut-off to 0.8, we saw that false negatives have also reduced. If it's necessary, we can call these false negatives. We could observe that by increasing the cutoff, there was a decrease in the number of false positives from 533 (in Q3) to 47 which was very less. So, we can increase the cut-offs on the predicted probabilities and save the resources with minimum callings. Also, the identified leads can be communicated with automated emails which will avoid unnecessary callings.