

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

ANSWERS - 1

- ☐ **Top three variables in our model which are contributing most towards the probability of lead getting converted are**
- ☐ **Do not Email – The leads which are showing the sub category value of do not call are more prone to go towards junk lead and hence this particular is a good indicator whether the lead will convert or not**
- ☐ **Total Time Spent on Website – Total Time Spent on spent on website is perhaps the most important numerical feature among all the features. Higher total time spent is showing a higher chance of lead conversion. Since total time spent is an indicator whether the Customer is actually intended to buy the course or not.**
- ☐ **Lead Origin – It is also one of the most important columns among all our columns since , there is a huge disparity among the percentage of lead conversion compared to other Origin sources. Also this can also act as one of the recalibration opportunity for the marketing , lead generation team , where they can focus to increase the ad visibility or lead generation from certain sources and on the other hand completely ignoring some of the others**

ANSWER 2

- **From our WOE analysis and IV interpretation**
- **Handling free copies – It can be seen that adding a supplementary course is becoming a factor where it helps to make client confident in our services**

- Welingak website is one of the dummy variable showing a good correlation coefficient in positive directions. It basically indicates that lead are more likely to convert from this model
- Currently working professions should be focused more on since they are showing a positive correlation in our model

ANSWER 3

- In the first case scenario we can basically set the threshold probability to as low as possible. It has upside that our overall positive and true positive rate will rise and can become close to 1, Where there will not be misclassification with regards to positive leads classified as junk leads. But this will create a counter effect by producing large number of leads which are false positive in nature. During this case the main agenda could be to tune the model to reach higher Sensitivity and reduce the Specificity to as low as possible. But it should be taken care of that the tradeoff should not be made to such an extent that True False Rate becomes close to zero or zero.

ANSWER 4

- During this time A higher Specificity could be achieved by increasing the probability but in this case it is recommended to reach for an harmonic mean value using f1 , to have a balanced trade off.