

Case Study – Lead Scoring – Subjective Questions

Submitted by: Sangeeta Kalra & Vikram Mathur

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

A1. The top three variables in the model are determined by reviewing the coefficients of the features. Based on the following three variables contribute the most towards the probability of the lead getting converted:

- i. Total Time Spent on Website
- ii. Lead Origin_Lead Add Form
- iii. 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?

A2. The top three categorical/dummy which should be focused on most in order to increase the probability of lead conversion are:-

- i. Lead Origin_Lead Add Form
- ii. Current Occupation_Working Professional
- iii. Do Not Email_True (Negative Correlation) OR Do Not Email_False (Positive Correlation)

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.

A3.

- i. The trained model should have a very high “True Positive Rate”. A value of 90% plus would be ideal. This value gives you the number of positives correctly predicted divided by the total number of positives. These metrics help ensure that the conversion (true positive) has higher probability.
- ii. Maximizing AUC ROC: The Area Under the Curve of the ROC curve (AUC ROC) gives an indication of how well the leads are ranked. A perfect model will have AUC ROC = 1. Hence, our strategy should be to increase the AUC ROC as much in order to get more success.
- iii. Another way to look at it, is that in the AUC ROC curve we can consider only the highest percentage of leads which were predicted to convert. This will specially reduce the model’s False Positive Rate (FPR).
- iv. Additionally, the Threshold value (cut-off value) can be tweaked, typically an increase in it could increase the TPR.

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.

A4.

- i. Here for the company it is not absolutely important for them to get a conversion since the target is achieved, but what is important is that a potential lead which in actual is a Yes (That is, is a True Positive) should not be missed out.
- ii. The strategy should be to build a model that has a higher Recall value so that we definitely capture the positive cases and they are not missed out.