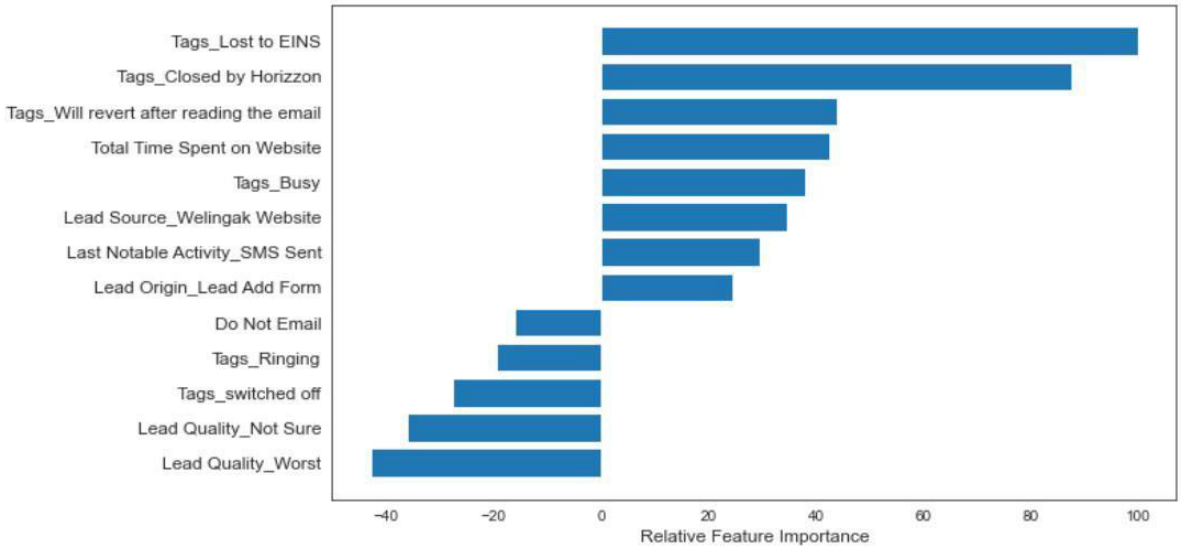
Leads Scoring Case Study Assignment Question & Answers

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

Ans : Following is the graph of relative importance of different features based on their coefficient values in the model:



From the graph, it's evident that “Tags_Lost to EINS”, “Tags_Closed by Horizzon”, and “Tags_Will” revert after reading the email are the top three variables contributing most significantly to the probability of lead conversion. These features, derived from the categorical variable Tags, show a positive correlation with lead conversion. Therefore, prioritizing leads with these tags can potentially enhance the conversion rate and improve overall business outcomes.

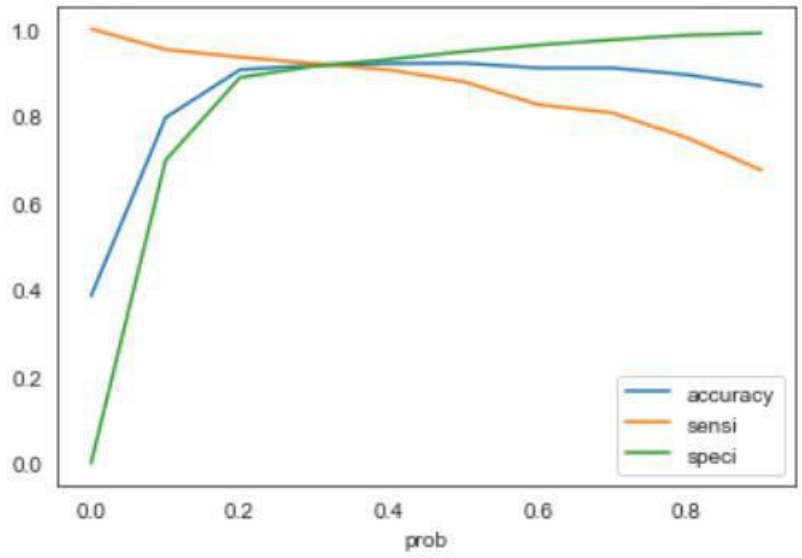
2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

Ans: From the graph above, 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 are:

I. Tags_Lost to EINS
II. Tags_Closed by Horizzon
III. Tags_Will revert after reading the email Answer to both the questions is same because the top 3 variables in the model are all categorical/dummy variables

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.

Ans : Here, the concept of sensitivity is required. $\boldsymbol{Sensitivity} = \boldsymbol{True\ Positives}/(\boldsymbol{True\ Positives} + \boldsymbol{False\ Negatives})$ 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 of sensitivity can be achieved for the model by changing the cutoff threshold for probability of lead conversion. For our model, below is the graph showing changes in Sensitivity, Specificity and Accuracy with change in the threshold:



As we can see, 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.

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.

Ans: To minimize the rate of useless phone calls during a period when the company has already met its quarterly target, it's essential to focus on specificity in our predictive model. Specificity measures the ability of the model to correctly identify non-converting leads out of all actual non-converting leads.

Since the company wants to avoid unnecessary phone calls, opting for high specificity is crucial. High specificity ensures that the model accurately predicts leads who are unlikely to convert, reducing the chances of making unnecessary calls. Although this approach may result in some missed conversions being classified as non-conversions, it aligns with the company's goal of minimizing phone calls during this period.

To achieve high specificity, we should choose a high threshold value for our predictive model. This decision will ensure that phone calls are primarily made to leads with a very high probability of conversion, optimizing the sales team's efforts and maximizing efficiency during this phase.