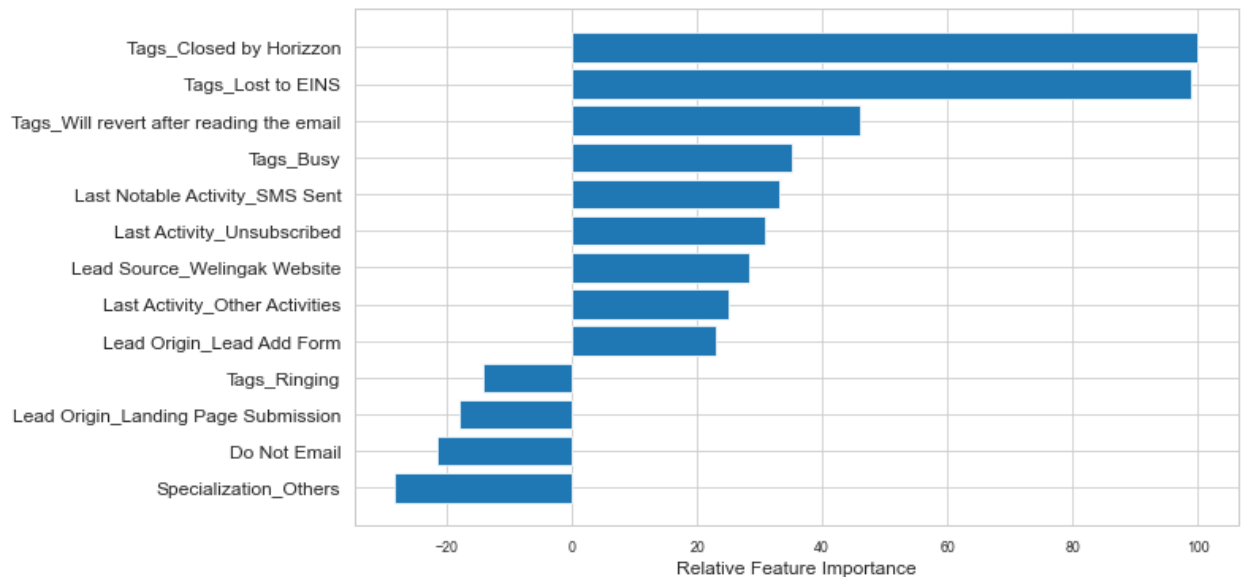


# Lead Scoring Assignment – Subjective Questions

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 which contribute most towards the probability of a lead getting converted are:

1. Tags\_Closed by Horizon
2. Tags\_Lost to EINS
3. Tags\_Will revert after reading the email

They contribute positively towards the probability of a lead conversion. Hence the company should focus more on the leads with these three tags.

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?

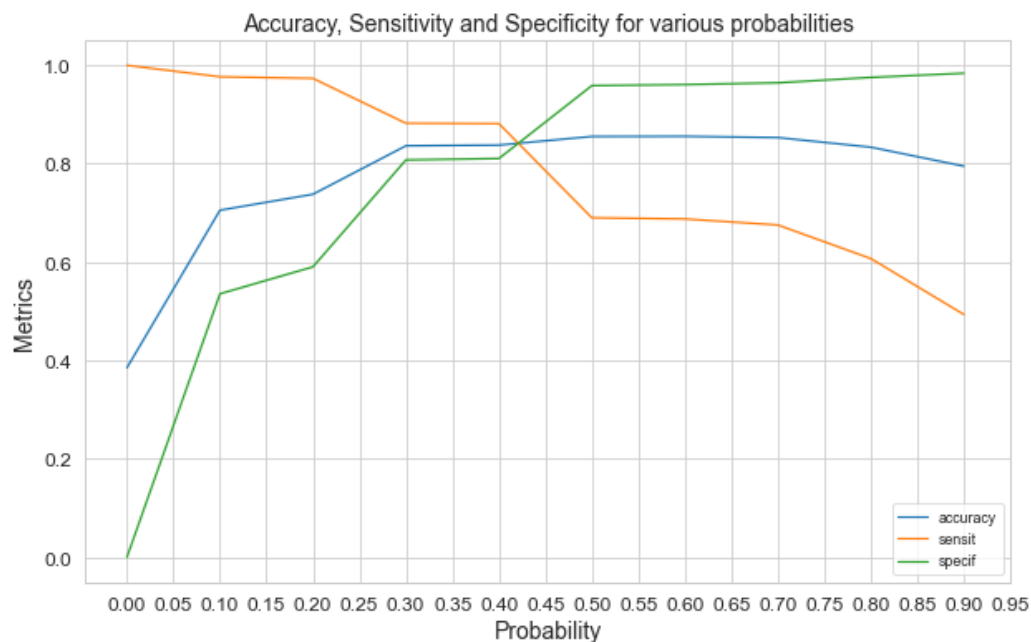
1. Tags\_Closed by Horizon
2. Tags\_Lost to EINS
3. Tags\_Will revert after reading the email

Again these are the top 3 dummy/categorical variables in the model that 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.

We can apply Sensitivity concept here. Sensitivity is defined as the number of actual conversions predicted correctly out of total number of actual conversions.

$$\text{Sensitivity} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})}$$



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.

Here, the concept of specificity can be applied.

$$\text{Specificity} = \text{True Negatives} / (\text{True Negatives} + \text{False Positives})$$

Specificity can be defined as the number of actual non-conversions predicted correctly out of total number of actual non-conversions.

From the above graph, we can see that the specificity increases as the threshold increases. In the given situation, we'll need a **high specificity** because high specificity will mean that our model will correctly predict almost all leads who are not likely to convert.

At the same time, it may misclassify some of the conversions as non-conversions. But as the company has already reached its target for a quarter and doesn't want to make phone calls unless it's extremely necessary, it is a good strategy to go for high specificity. It will ensure that the phone calls are only made to customers who have a very high probability of conversion. To achieve high specificity, we need to choose a high threshold value.