

Assignment Subjective Questions - Solutions

- Praveen Sankadal

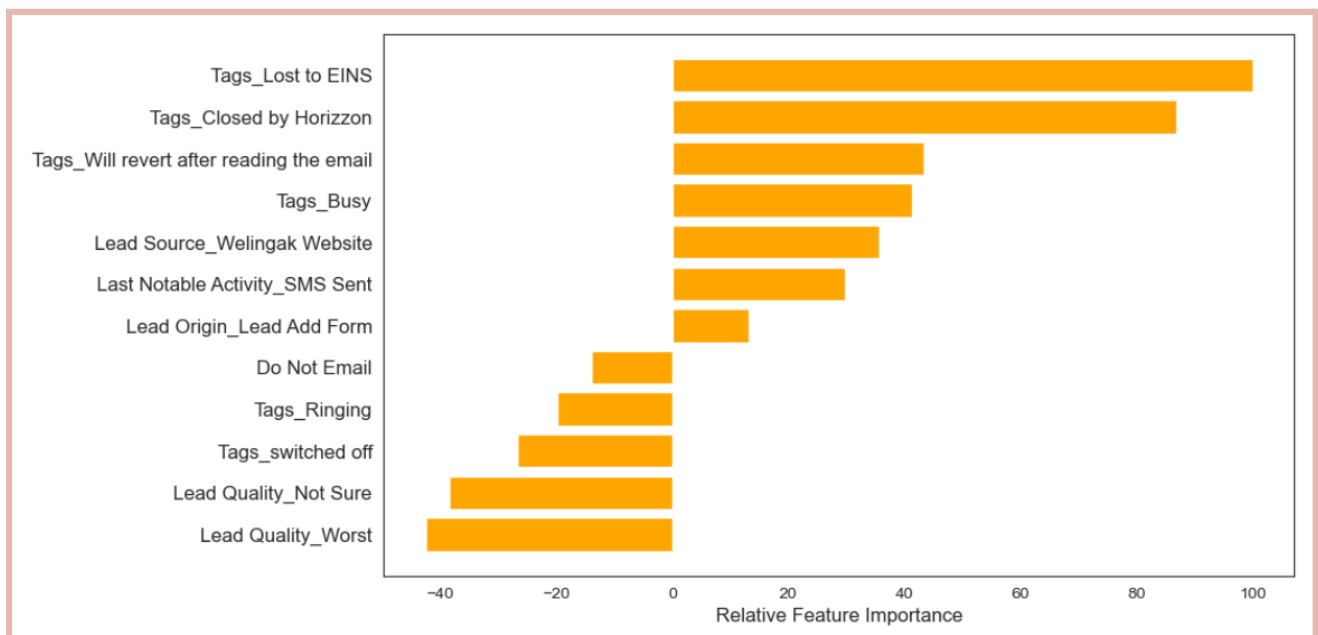
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Ans:

top three variables with the most significant impact on the probability of lead conversion, listed in decreasing order, are:

- Tags_Lost to EINS
- Tags_Closed by Horizzon
- Tags_Will revert after reading the email

These features are generated as dummy variables from the categorical variable "Tags" and contribute positively to the likelihood of lead conversion. The findings suggest that the company should prioritize leads associated with these three specific 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?

Ans:

The three most crucial categorical/dummy variables in the model that demand the highest attention for boosting the likelihood of lead conversion are:

- Tags_Lost to EINS
- Tags_Closed by Horizzon
- Tags_Will revert after reading the email

Both questions share the same answer, as the top three 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.

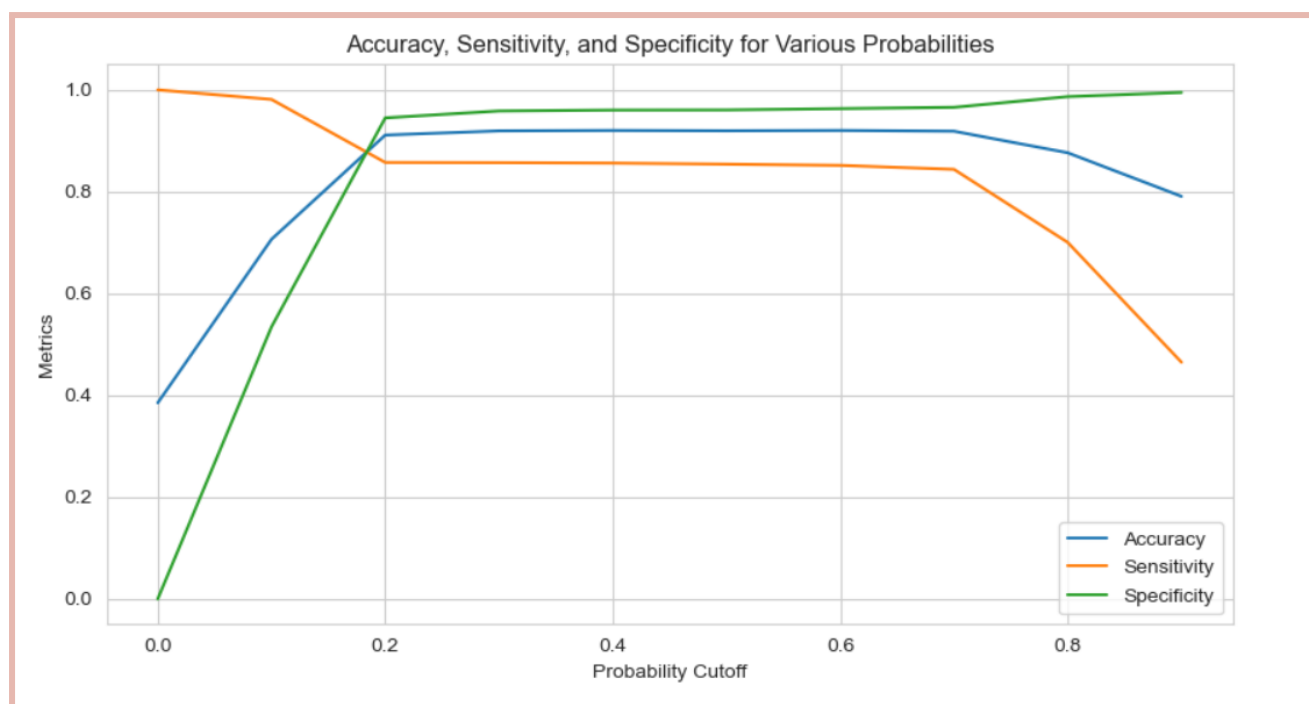
Ans:

Here we use the concept of Sensitivity,

Sensitivity measures the ability of a model to correctly identify positive cases, specifically the number of actual conversions predicted correctly out of the total actual conversions.

$$\text{Sensitivity} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})}$$

Changing the threshold for the probability of lead conversion can result in different sensitivity values for the model.



In our model, as the threshold increases, sensitivity decreases. For our situation, high sensitivity is crucial because it ensures that the model correctly predicts most leads likely to convert, aligning with the goal of making lead conversion more aggressive.

Choosing a low threshold value helps achieve high sensitivity, making it a suitable strategy given the company's current approach.

- 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 address this question, we follow a similar approach as the previous one, focusing on the concept of specificity.

Specificity is defined as the number of actual non-conversions predicted correctly out of the total number of actual non-conversions. It can be calculated using the formula:

$$\text{Specificity} = \frac{\text{True Negative}}{(\text{True Negative} + \text{False Positive})}$$

In our model, we observe that specificity increases as the threshold increases. In our situation, aiming for high specificity is essential because it ensures that the model correctly predicts almost all leads not likely to convert. This strategy aligns with the company's goal, as it has already met its quarterly target and wants to avoid unnecessary phone calls.

Although there might be misclassifications of conversions as non-conversions, the priority is to focus on high specificity. This ensures that phone calls are directed only to customers with a very high probability of conversion. Achieving high specificity requires choosing a high threshold value.