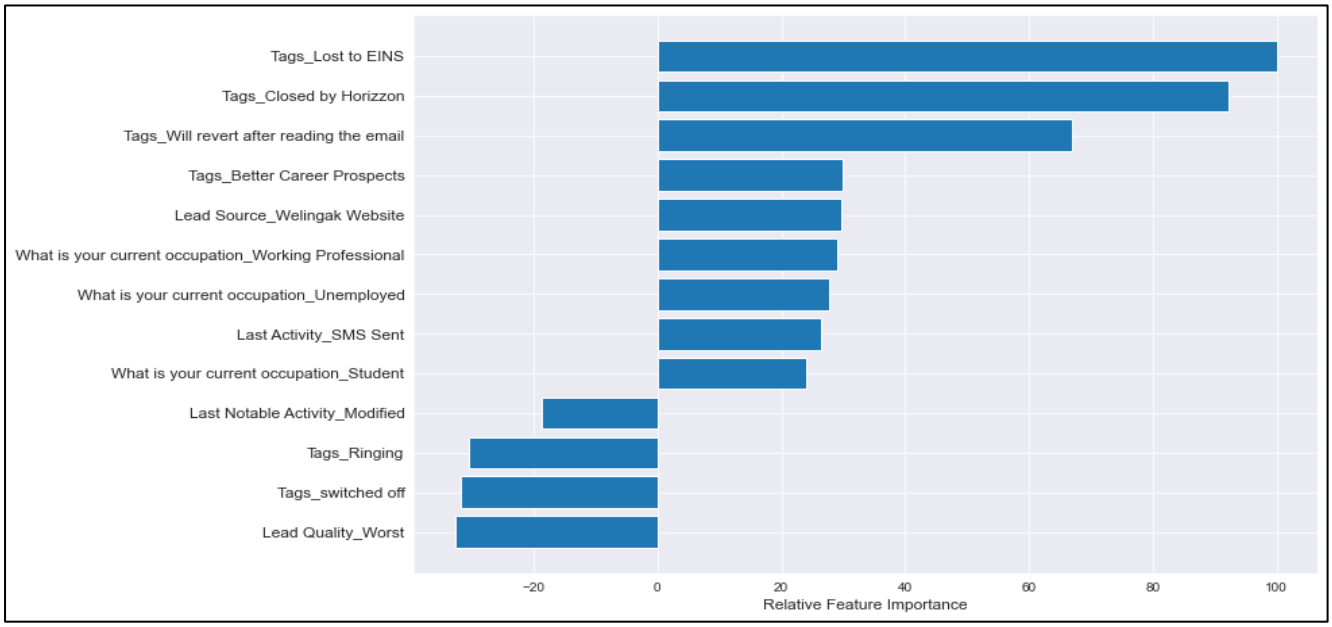
Assignment Subjective Questions

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

- The following graph displays the relative importance of various features based on their coefficient values in the model:



Referring to the graph, the top three variables contributing most significantly to the probability of lead conversion in decreasing order of impact are:

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

All of these features are dummy variables derived from the categorical variable "Tags." These features have a positive influence on the probability of lead conversion. These findings suggest that the company should prioritize leads with these three tags, as they have a significant impact on conversion rates.

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?**

Based on the graph provided, the top 3 categorical/dummy variables in the model that should receive the highest focus to enhance the probability of lead conversion are:

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

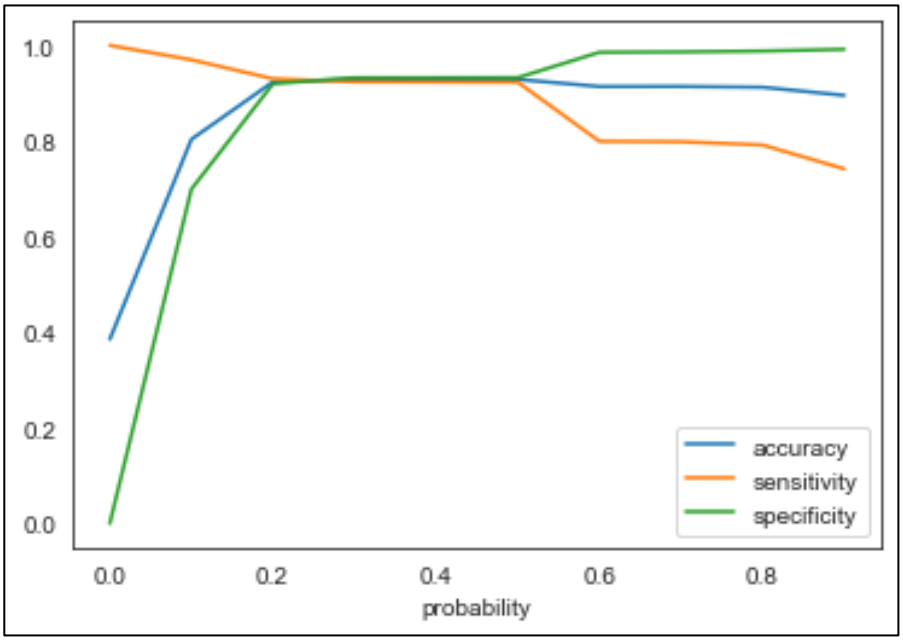
Indeed, the answers to both questions are the same because the top 3 variables in the model are all categorical/dummy variables, which should be the primary focus for increasing the probability of lead conversion.

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.**

- Certainly, the concept of sensitivity is crucial in this context, as it helps identify the key variables that have the most significant impact on increasing the probability of lead conversion.

$$\boldsymbol{Sensitivity = True\ Positives / (True\ Positives + False\ Negatives)}$$

- In the context of our model, sensitivity can be defined as the proportion of actual conversions that are accurately predicted out of the total number of actual conversions.
- Various sensitivity values can be attained for the model by altering the cutoff threshold for the probability of lead conversion.
- For our model, the following graph illustrates how Sensitivity, Specificity, and Accuracy change with alterations in the threshold:



- As evident from the graph, sensitivity declines as the threshold increases. In the given scenario, a high sensitivity is crucial because it ensures that our model accurately predicts almost all leads likely to convert, aligning with the company's goal of capturing as many potential conversions as possible.
- While pursuing high sensitivity, it's important to keep in mind that this approach might lead to some overestimation and the misclassification of non-conversions as conversions. However, given the company's temporary surplus of manpower for two months and its aim to intensify lead conversion efforts by reaching out to as many potential leads as possible, selecting a low threshold value is indeed a prudent strategy. This would prioritize identifying most potential conversions, even at the expense of a few false positives.

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.**

- The approach to addressing this question is similar to the previous one.
- Certainly, the concept of specificity is required in this context

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

- In the context of our model, specificity can be defined as the proportion of actual non-conversions that are accurately predicted out of the total number of actual non-conversions.
- As observed in the previous graph, having high specificity means our model will accurately identify leads that are unlikely to convert, reducing unnecessary efforts. This approach is suitable when the company wants to conserve resources and concentrate on the most promising leads.
- While focusing on high specificity, it's necessary to understand that this approach may result in misclassification of some actual conversions as non-conversions. However, considering that the company has already achieved its quarterly target and aims to minimise phone calls unless absolutely necessary, choosing high specificity is advisable strategy. This helps in avoiding unnecessary efforts and conserving resources, even though it may potentially miss a few genuine conversion opportunities.
- Selecting a high specificity approach ensures that phone calls are primarily directed towards customers with a very high likelihood of conversion, which optimizes resource utilization. Achieving high specificity requires choosing a high threshold value, which minimises the chances of false positives and focuses efforts on the most promising leads.