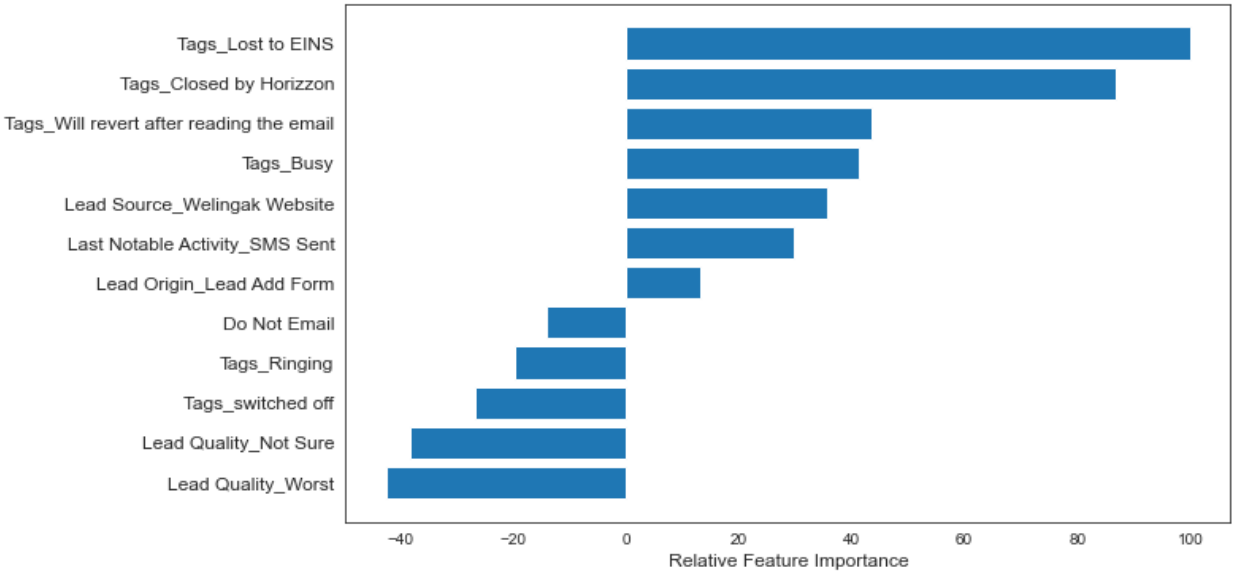
Answers to Assignment Subjective Questions

- Priyanka Rotte

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

Answer:

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



From the graph, the three variables which contribute most towards the probability of a lead conversion in decreasing order of impact are:

I. **Tags_Lost to EINS**
II. **Tags_Closed by Horizzon**
III. **Tags_Will revert after reading the email**

All these features are dummy features created from the categorical variable Tags. These features **contribute positively** towards the probability of a lead conversion. These results indicate that 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?

Answer:

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.

Answer:

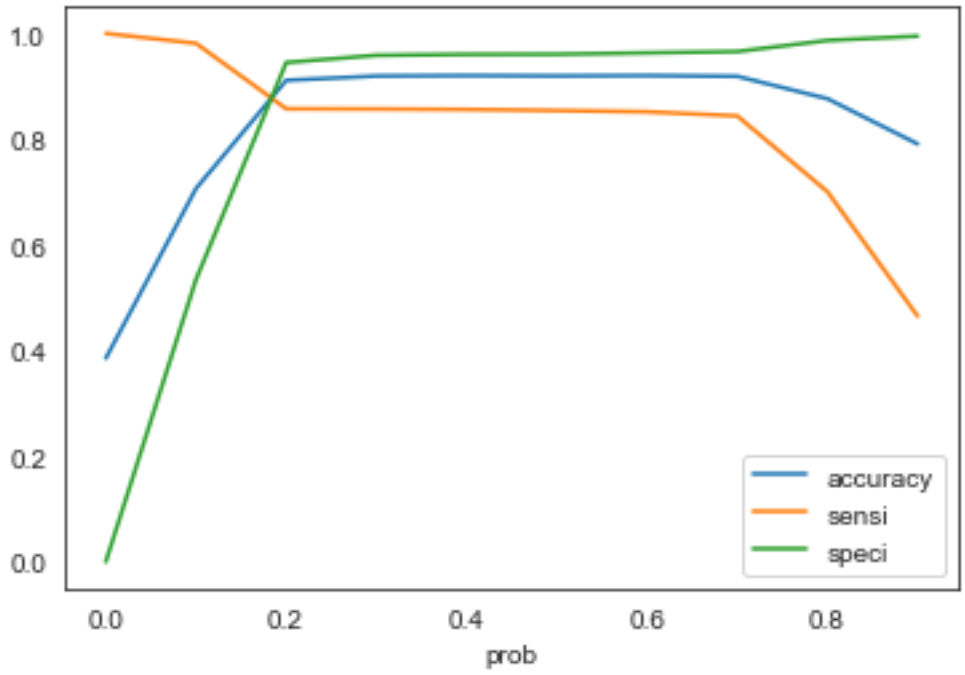
Here, the concept of sensitivity is required.

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

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.

Answer:

The approach to answer this question is similar to the last one.
Here, the concept of specificity is required.

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

With respect to our model, 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**.
