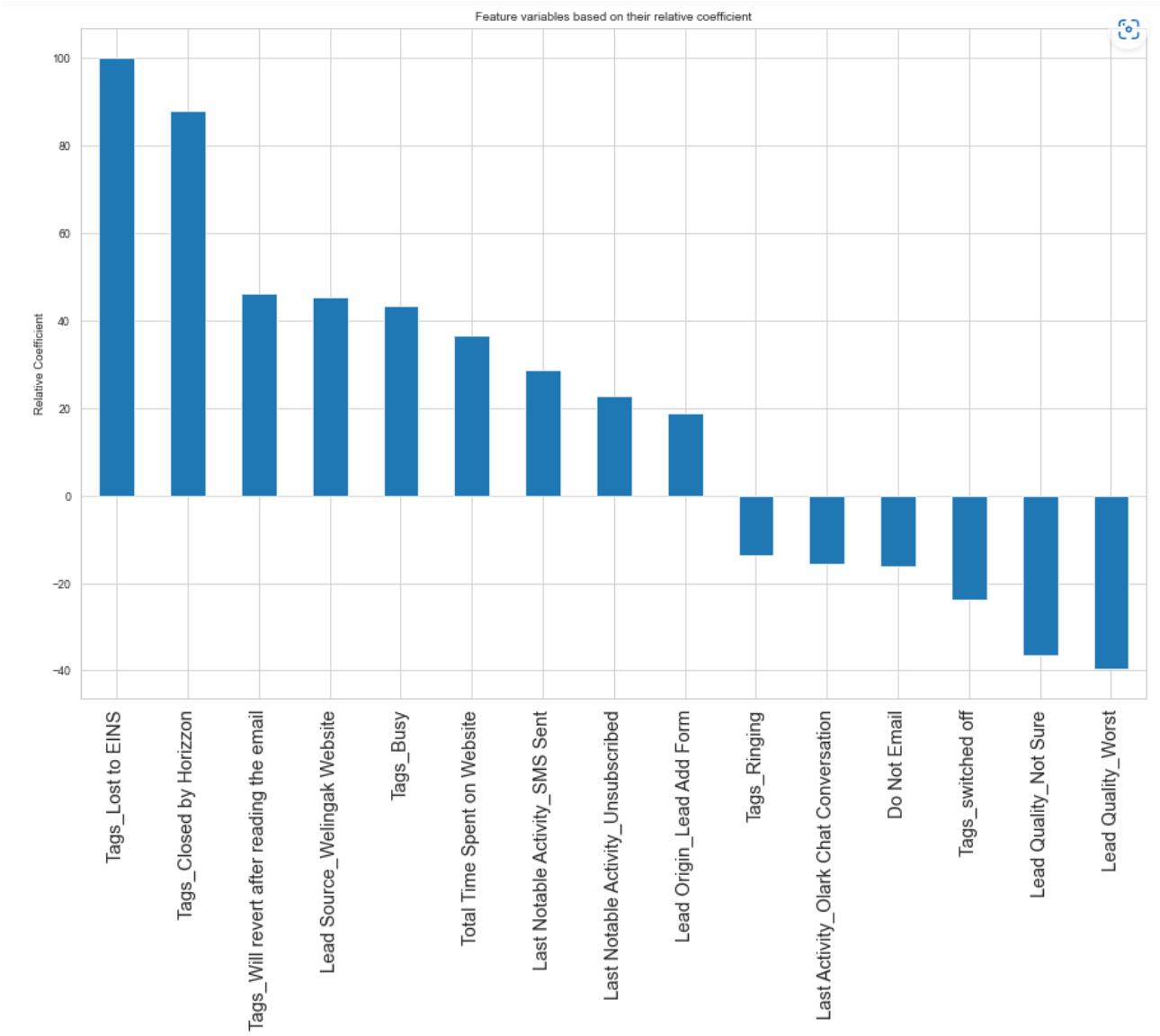
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Ans 1 : - As per the below graph, we have analyzed the relative importance of different features based on their coefficient values in the model.



As per the above graph, the top three variables which contribute most toward the probability of lead conversion in decreasing order of impact are:

I. Tags_Lost to EINS (Coeff= 8.7797)
II. Tags_Closed by Horizzon (Coeff= 7.7226)

III. Tags_Will revert after reading the email (Coeff= 4.0649)

All these features are dummy features created from the categorical variable Tags. These features contribute positively towards the probability of 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?

Ans 2 : - 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 (Coeff= 8.7797)
II. Tags_Closed by Horizzon (Coeff= 7.7226)
III. Tags_Will revert after reading the email (Coeff= 4.0649)

The answer to both questions is the 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 many of such people as possible. Suggest a good strategy they should employ at this stage.

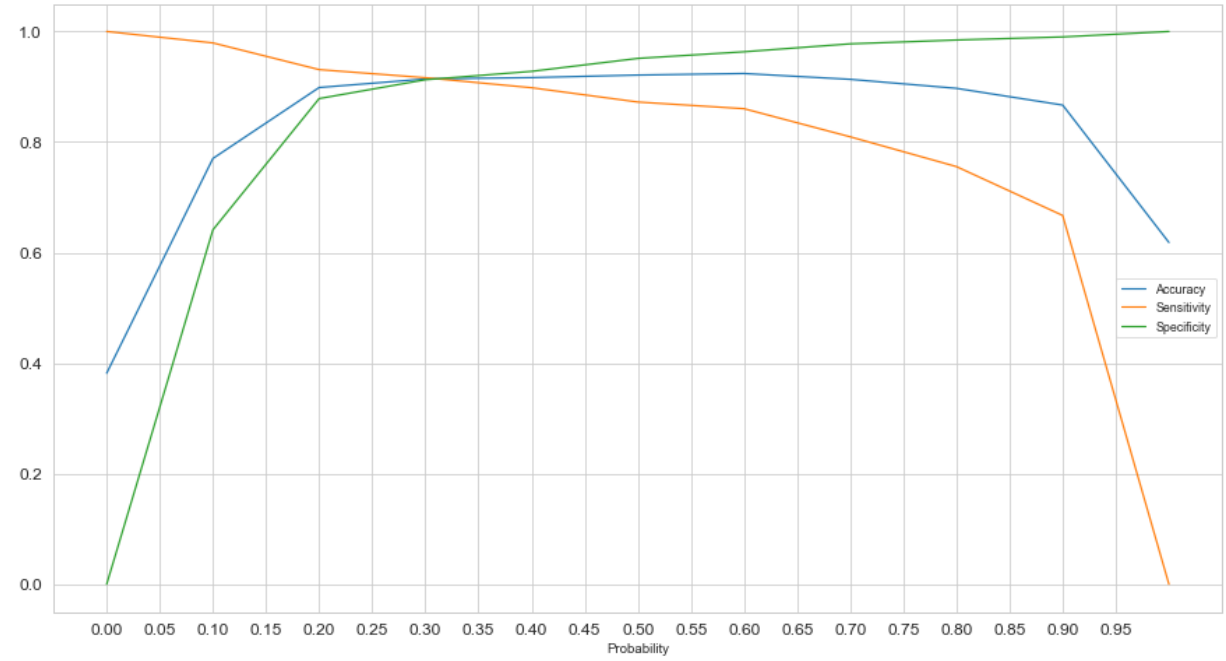
Ans 3 : - X Education have 2 months where they have extra team size where they can cover maximum Potential leads predicted through the model and tried to reach.

So as per our understanding related to the model, we need to Increase **Sensitivity** means our model will identify all leads which are most likely to convert into potential customers.

we can choose a lower threshold value for Conversion Probability. This helps to increase in identification of all leads which will more likely to convert. And X Education can utilize more new extra sales interns to make phone calls to all possible leads.

When the probability thresholds are very low, the sensitivity is very high and the specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.

The below graph shows how the Sensitivity and Specificity rating changes with a change in the threshold value for our model: -



Also, **after analyzing the results of EDA our new Interns must approach people by** checking categories: -

- They spend a lot of time on the website and this can be done by making the website interesting and thus bringing them back to the site.
- People having Tags “Will revert after reading emails” can be possible targeted leads
- Their last activity is through SMS & Email Opened can be targeted.
- They are working professionals.

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 4 : - The approach to answering this question is similar to the previous answer: -

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 the 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.

As a result, the Sales team won't have to make unnecessary phone calls and can focus on some new work.