

LOGISTIC REGRESSION ASSIGNMENT

LEAD SCORING ASSIGNMENT **SUBJECTIVE QUESTION ANSWER**

BY

Monalisa Banerjee

Swetha Mennde

Amit kumar

|

Student of IIIT - Bangalore and UpGrad

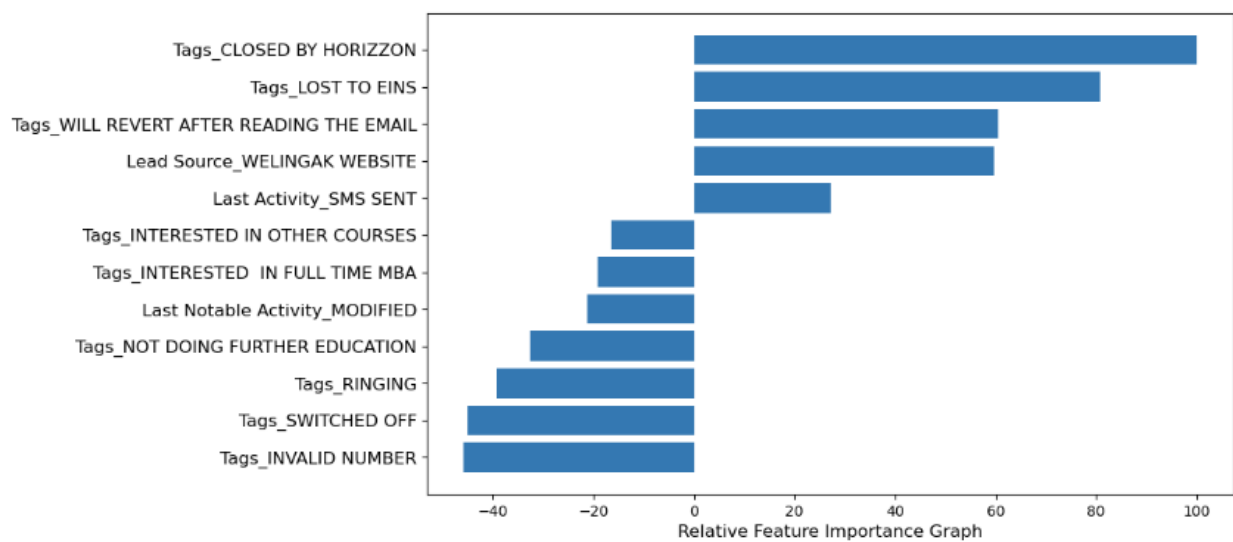
Batch : DS-C70

|

Date : 20-01-2025

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

Ans: - After trying different variations of the Logistic Regression model we finally reached to the Accuracy of almost 90%. Considering this as final Model we proceed with further analysis. From the final model we received below Feature Importance Graph we are seeing below,

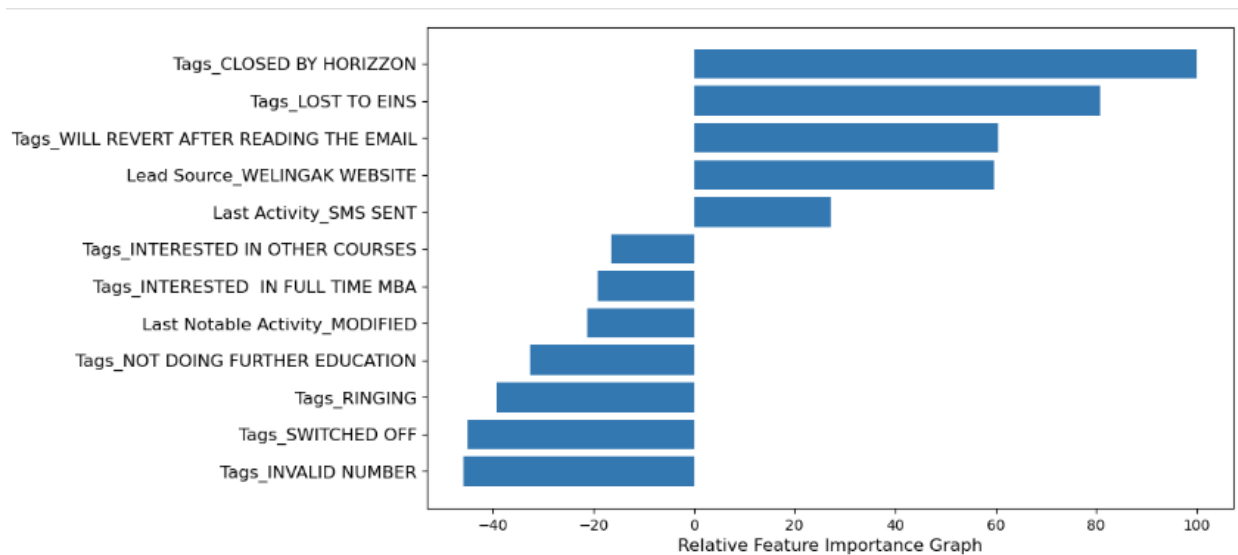


From the above graph its very much evident that we got 3 best features are

- i. Tags
- ii. Lead Source
- iii. Last Activity

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: - As per the graph below,



We have got 3 best categorical/dummy variables with Positive coefficient as below,

- i. Tags_CLOSED BY HORIZZON
- ii. Tags_LOST TO EINS
- iii. Tags_WILL REVERT AFTER READING THE EMAIL

As the priority is how to increase potential Leads, so considering positive coefficient for this case.

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: - Sensitivity tells us how well our model avoids missing actual conversions. A high sensitivity means the model is good at identifying potential customers.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

TP refers to, **True Positives**: The model correctly predicted a conversion.

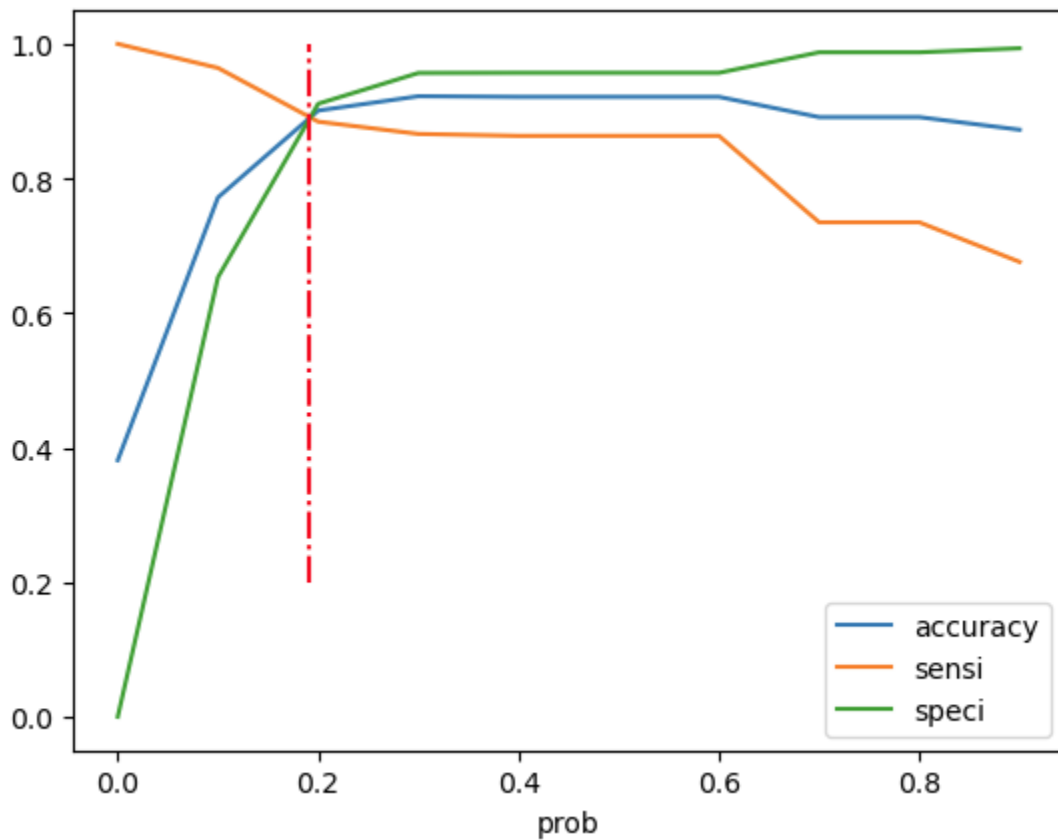
FN refers to, **False Negatives**: The model incorrectly predicted no conversion when there actually was one.

If we change the probability of the Target Variable i.e Converted, then we will be able to see different values of the sensitivity for the final model.

We can choose the cutoff from these different sensitivity values and accordingly we can go ahead with the deciding the final features and exactly these features will eventually help Interns to get higher conversion rate. So, these features will be the final predicted features where the Interns will focus to invest more time and patience.

Besides the Interns must ignore the features which are coming with Negative coefficients as these features has least probability of getting converted.

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 probability 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.

Ans: - The approach to answer this question is exactly opposite to the previous one.

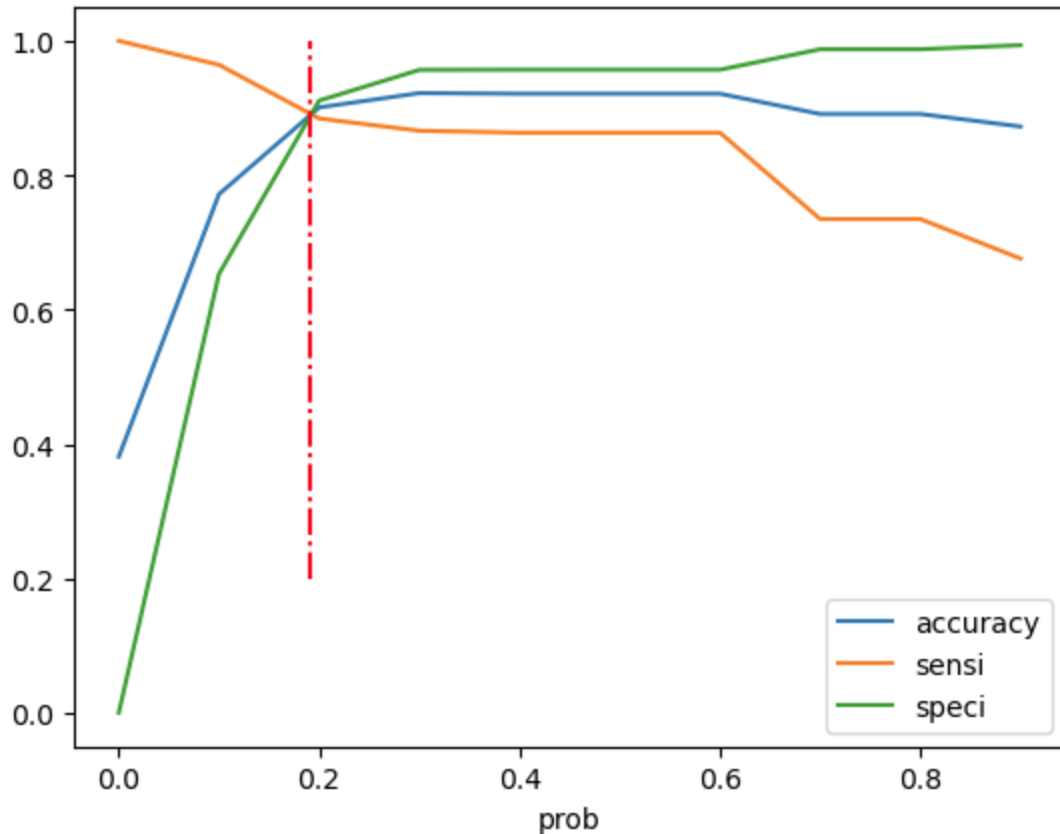
In the last strategy we used sensitivity to increase the conversion rate. Now as the target has been achieved, and the man-power is also going to less during this time so, the concept of specificity will be applicable here.

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

TN refers to, **True Negatives**: The model correctly predicted a lead will not be converted.

FP refers to, **False Positives**: The model incorrectly predicted conversion when there actually was no conversion.

From the below graph, we can see that the specificity increases as the threshold increases.



In

In the given situation, we'll use a high specificity because high specificity means that the model will correctly predict almost all leads which are not likely to convert.

At the same time, it may misclassify some of the conversions as non-conversions. But that's okay as the company has already reached its target for a quarter and doesn't want to make phone calls unless it's extremely urgent.

Hence, it is a good strategy to go with 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 probability value.