

## LEAD SCORING SUBJECTIVE ASSIGNMENT

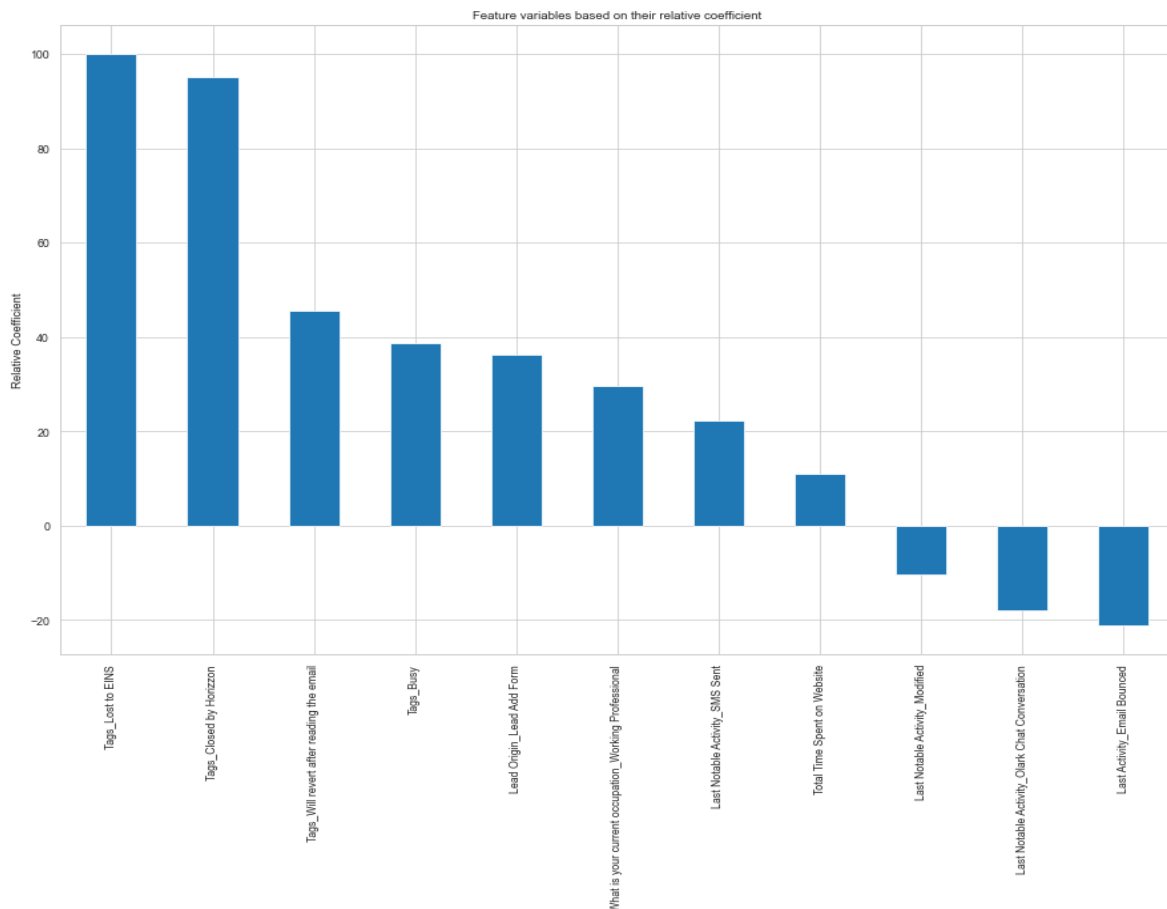
SUBMITTED BY: GURPREET KAUR ||| SHIVAM SHARMA ||| (DSC43 BATCH)

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

**Top 3 variables that Contributing Most towards the Probability of Lead conversion in decreasing order of impact are:**

1. Tags\_Lost to EINS: Leads that have been tagged as 'Lost to EINS' also contribute to the conversion to a considerable extent. **(Relative Coefficient Value: 100.00)**
2. Tags\_Closed by Horizzon: Leads that have been assigned Tags as 'closed by horizon' have the highest probability of conversion. **(Relative Coefficient Value: 95.16)**
3. Tag\_Will revert after reading the email: Leads that have been tagged as 'will revert after reading the mail' also have significant correlation with the conversion. **(Relative Coefficient Value: 45.51)**

Following is the graph of Relative Importance of Different Features based on their Coefficient Values in the Model:



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?

Tags_Lost to EINS	100.00
Tags_Closed by Horizzon	95.16
Tags_Will revert after reading the email	45.51
Tags_Busy	38.72
Lead Origin_Lead Add Form	36.18
What is your current occupation_Working Professional	29.60
Last Notable Activity_SMS Sent	22.22
Total Time Spent on Website	10.96
Last Notable Activity_Modified	-10.47
Last Notable Activity_Olark Chat Conversation	-17.97
Last Activity_Email Bounced	-21.22
dtype: float64	

Again, based on the **Coefficient Values** from the screen shot above, the following are the **Top Three Categorical/Dummy Variables** that should be focused the most in order to increase the probability of Lead Conversion:

- Tags** – (Lost to EINS, Closed by Horizzon, Will revert after reading the email)
  - Lead Origin** – (Lead Add Form)
  - What is your current occupation** – (Working Professional)
- 

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.

After analyzing the results of EDA our new Interns must approach people by Phone calls by checking categories:

- Company should make calls to the **“Working Professionals”** as they are more likely to get converted.
- Who visits websites repeatedly** or **Who spend much time on website** and this can be done by making website easier and more informative.
- Their **Last Activity** is through SMS & Email Opened can be targeted.
- People having **Tags “Will revert after reading emails”** can be possible targeted leads
- Last Notable Activity\_Had a Phone Conversation**

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.

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

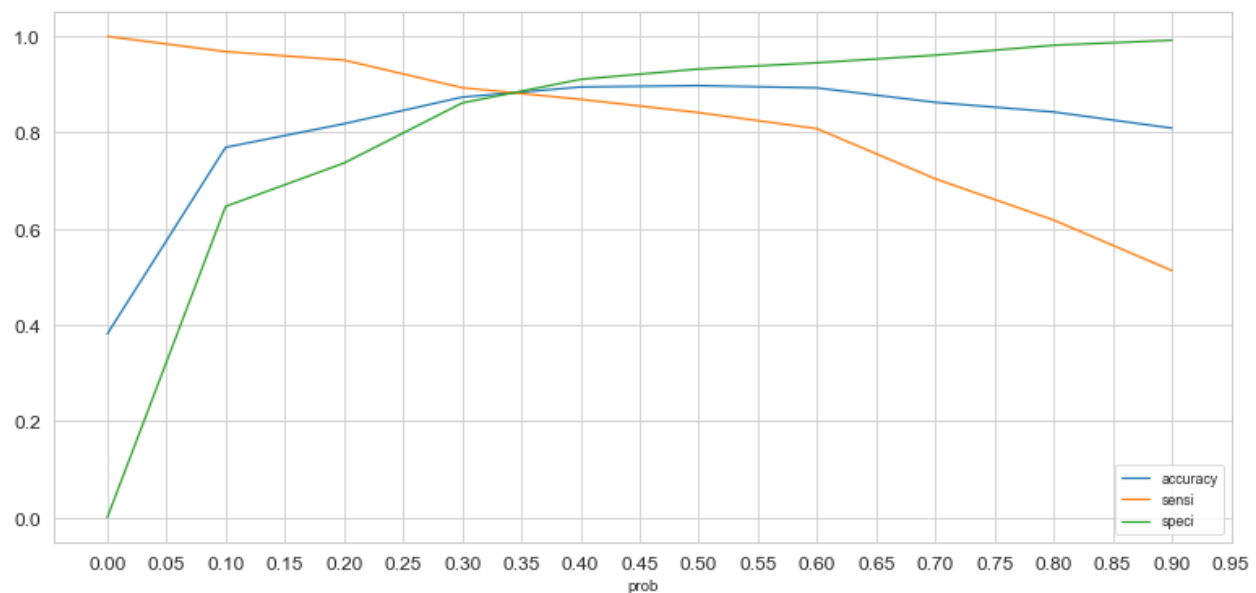
With respect to our model, **Sensitivity** can be defined as the number of actual conversions predicted correctly out of total number of actual conversions. To achieve high sensitivity, we need to choose a low threshold value.

Similarly, Specificity can be defined as the ratio of total no of actual non-Conversions correctly predicted to the total number of actual non-Conversions.

**High sensitivity** implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the **Conversion likelihood**, i.e. it will misclassify some non-Conversion cases as Conversions. Now, since **X Education has more man-power** for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all of the potential leads, we can choose a **lower threshold value for Conversion Probability**.

This will ensure the Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible. 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: **(0.34 is the Optimum Point to take it as a cutoff probability)**



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.

- They **need to focus on SMS and automated emails instead of phone call** unless its emergency.
- The above strategy can be used but with the customers that have a very high chance of buying the course.
- Leads who have **applied for 'Do Not Email' already does not needs** to be attended again.
- Also **based on the previous chat conversations** if the lead is classified as '**Might be**' or '**Worst**' then those leads can be **Ignored**.

- e. Leads who shared their contact number in the website or through email but either are '**Busy**' or '**Ringling**' i.e., **not answering** to the calls can also be **Ignored** as they are less likely to get converted.
- f. **Prioritizing** can be done on the **basis of lead score**. Leads that have **more than 80%** of lead score can be **Targeted**.

Here, the concept of **Specificity** is required.

***Specificity = True Negatives / (True Negatives + 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.**

**THANK YOU**