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

Answer

1. **Tags_Lost to EINS**
2. **Tags_Closed by Horizzon**
3. **Tags_Will revert after reading the email**

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:

1. **Tags_Lost to EINS**
2. **Tags_Closed by Horizzon**
3. **Tags_Will revert after reading the email**

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:

High sensitivity implies that our model will correctly identify almost all leads who are likely to convert.

Since there is 2 months of high man-power and they wish to make the lead conversion more aggressive, we can choose a high sensitivity and closer specificity to ensure that the conversion rate increases.

From below picture of sensitivity, specificity and accuracy probability we can say that a probability of 0.3 best suits the purpose of the company. Highlighted below are the details.

```
num = [0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9]
for i in num:
    cm1 = metrics.confusion_matrix(y_train_pred_final.Converted, y_train_pred_final[i] )
    total1=sum(sum(cm1))
    accuracy = (cm1[0,0]+cm1[1,1])/total1

    speci = cm1[0,0]/(cm1[0,0]+cm1[0,1])
    sensi = cm1[1,1]/(cm1[1,0]+cm1[1,1])
    cutoff_df.loc[i] =[ i ,accuracy,sensi,speci]
print(cutoff_df)
```

	prob	accuracy	sensi	speci
0.0	0.0	0.385136	1.000000	0.000000
0.1	0.1	0.705086	0.981603	0.531882
0.2	0.2	0.909148	0.860589	0.939565
0.3	0.3	0.920013	0.859771	0.957746
0.4	0.4	0.919853	0.858953	0.958005
0.5	0.5	0.919383	0.851594	0.961844
0.6	0.6	0.920170	0.851594	0.963124
0.7	0.7	0.919383	0.845462	0.965685
0.8	0.8	0.878917	0.706868	0.986684

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:

Based on the same logic as applied above, we will now try to see the high specificity but having low sensitivity as the strategy is to go slow on conversion since the target is met.

This means that the sensitivity is low i.e. actual leads called but since specificity is high then the conversion rate is high with minimal effort.

Below highlighted shows the best possible results for this.

```
num = [0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9]
for i in num:
    cm1 = metrics.confusion_matrix(y_train_pred_final.Converted, y_train_pred_final[i] )
    total1=sum(sum(cm1))
    accuracy = (cm1[0,0]+cm1[1,1])/total1

    speci = cm1[0,0]/(cm1[0,0]+cm1[0,1])
    sensi = cm1[1,1]/(cm1[1,0]+cm1[1,1])
    cutoff_df.loc[i] =[ i ,accuracy,sensi,speci]
print(cutoff_df)
```

	prob	accuracy	sensi	speci
0.0	0.0	0.385136	1.000000	0.000000
0.1	0.1	0.705086	0.981603	0.531882
0.2	0.2	0.909148	0.860589	0.939565
0.3	0.3	0.920013	0.859771	0.957746
0.4	0.4	0.919855	0.858953	0.958003
0.5	0.5	0.919383	0.851594	0.961844
0.6	0.6	0.920170	0.851594	0.963124
0.7	0.7	0.919383	0.845462	0.965685
0.8	0.8	0.878917	0.706868	0.986684