

SUBJECTIVE QUESTION

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

The coefficients of these influential variables are as follows:

- a. 'Tags_Lost to EINS' (Coefficient factor = 8.745353)
- b. 'Tags_Closed by Horizon' (Coefficient factor = 8.753978)
- c. 'Lead Quality_Worst' (Coefficient factor = -3.539447) The model achieved a remarkable Sensitivity of 1, implying that it correctly predicts 100% of the converted customers (Positive conversions).

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?

In the ultimate model, the three most influential categorical/dummy variables are 'Tags_Lost to EINS', 'Tags_Closed by Horizon', and 'Lead Quality_Worst,' as determined by the absolute values of their coefficient factors.

'Tags_Lost to EINS' and 'Tags_Closed by Horizon' were derived from the original categorical variable 'Tags' through encoding, while 'Lead Quality_Worst' resulted from encoding the categorical variable 'Lead Quality.'

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.

```

1 leadScore_build_model_cutoff(X_leadScore_train[leadScore_column], y_leadScore_train, X_leadScore_test[leadScore_column],

```

-----Result of training data-----

| | Prospect ID | Converted | Convert_Probability | Convert_predicted | Lead_Score |
|---|-------------|-----------|---------------------|-------------------|------------|
| 0 | 2746 | 1 | 0.998158 | 1 | 100 |
| 1 | 6543 | 0 | 0.002923 | 0 | 0 |
| 2 | 7458 | 1 | 0.910774 | 1 | 91 |
| 3 | 7448 | 1 | 0.988297 | 1 | 99 |
| 4 | 6134 | 1 | 0.950594 | 1 | 95 |

-----Result of test data-----

| | Prospect ID | Converted | Convert_Probability | Convert_predicted | Lead_Score |
|---|-------------|-----------|---------------------|-------------------|------------|
| 0 | 141 | 0 | 1.0 | 1 | 100 |
| 1 | 7418 | 0 | 1.0 | 1 | 100 |
| 2 | 6408 | 1 | 1.0 | 1 | 100 |
| 3 | 3128 | 0 | 1.0 | 1 | 100 |
| 4 | 1490 | 1 | 1.0 | 1 | 100 |

-----Model Evaluation Metrics-----

lead Score Confusion Matrix :

```

[[ 6 1267]
 [ 0 761]]

```

lead score Accuracy : 0.37708947885939037
lead Score Sensitivity : 1.0
lead score Specificity : 0.004713275726630008
lead score Precision : 0.3752465483234714

| | Prospect ID | Converted | Convert_Probability | Convert_predicted | Lead_Score |
|------|-------------|-----------|---------------------|-------------------|------------|
| 0 | 141 | 0 | 1.0 | 1 | 100 |
| 1 | 7418 | 0 | 1.0 | 1 | 100 |
| 2 | 6408 | 1 | 1.0 | 1 | 100 |
| 3 | 3128 | 0 | 1.0 | 1 | 100 |
| 4 | 1490 | 1 | 1.0 | 1 | 100 |
| ... | ... | ... | ... | ... | ... |
| 2029 | 6261 | 0 | 1.0 | 1 | 100 |
| 2030 | 5868 | 1 | 1.0 | 1 | 100 |
| 2031 | 1180 | 1 | 1.0 | 1 | 100 |
| 2032 | 5656 | 1 | 1.0 | 1 | 100 |
| 2033 | 9119 | 0 | 1.0 | 1 | 100 |

2034 rows × 5 columns

- 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.

```
1 leadScore_build_model_cutoff(X_leadScore_train[leadScore_column], y_leadScore_train, X_leadScore_test[leadScore_column])
```

```
-----Result of training data-----
Prospect ID  Converted  Convert_Probability  Convert_predicted  Lead_Score
0      2746      1      0.998158      1      100
1      6543      0      0.002923      0      0
2      7458      1      0.910774      1      91
3      7448      1      0.988297      1      99
4      6134      1      0.950594      1      95
```

```
-----Result of test data-----
Prospect ID  Converted  Convert_Probability  Convert_predicted  Lead_Score
0      141      0      1.0      1      100
1      7418      0      1.0      1      100
2      6408      1      1.0      1      100
3      3128      0      1.0      1      100
4      1490      1      1.0      1      100
```

```
-----Model Evaluation Metrics-----
```

lead Score Confusion Matrix :

```
[[ 14 1259]
 [ 0  761]]
```

lead score Accuracy : 0.3810226155358899

lead Score Sensitivity : 1.0

lead score Specificity : 0.010997643362136685

lead score Precision : 0.37673267326732673

| | ProspectID | Converted | Convert_Probability | Convert_predicted | Lead_Score |
|------|------------|-----------|---------------------|-------------------|------------|
| 0 | 141 | 0 | 1.0 | 1 | 100 |
| 1 | 7418 | 0 | 1.0 | 1 | 100 |
| 2 | 6408 | 1 | 1.0 | 1 | 100 |
| 3 | 3128 | 0 | 1.0 | 1 | 100 |
| 4 | 1490 | 1 | 1.0 | 1 | 100 |
| ... | ... | ... | ... | ... | ... |
| 2029 | 6261 | 0 | 1.0 | 1 | 100 |
| 2030 | 5868 | 1 | 1.0 | 1 | 100 |
| 2031 | 1180 | 1 | 1.0 | 1 | 100 |
| 2032 | 5656 | 1 | 1.0 | 1 | 100 |
| 2033 | 9119 | 0 | 1.0 | 1 | 100 |

2034 rows × 5 columns