

Q1 . Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Answers –

The logistic regression model estimates the probability that the target variable takes on a specific value, rather than directly predicting the target variable itself. This probability is then compared to a predefined cutoff to determine the predicted value of the target variable.⁵

The logistic regression model is employed to estimate the probability of a customer converting. We have chosen an optimal cutoff of 0.27, meaning that any lead with a probability greater than 0.27 is classified as a Hot Lead (likely to convert), while any lead with a probability of 0.27 or less is classified as a Cold Lead (unlikely to convert).

Our final Logistic Regression Model is constructed using 15 features:

- Features used in final model are ['Total Time Spent on Website', 'What is your current occupation', 'Lead Quality', '', 'Tags', 'Lead Source', 'Page view Per Visit', 'What matters most to you in choosing a course', 'Lead Origin', 'Do Not Email', 'Lead Profile', 'TotalVisits', 'Last Notable Activity', 'Through Recommendations', 'A free copy of MasteringThe Interview', 'Search', 'How did you hear about X Education', 'Do Not Call', 'Digital Advertisement', 'Specialization', 'Asymmetrique Activity Index', 'Last Activity', 'City', 'Magazine', 'Update me on Supply Chain Content', 'Get updates on DM Content', 'Receive More Updates About Our Courses', 'Newspaper', 'I agree to pay the amount through cheque', 'X Education Forums', 'Newspaper Article', 'Tags_switched off', 'Lead Quality_Not Sure', 'Lead Quality_Worst', 'Last Notable Activity_Modified', 'Last Notable Activity_Olark Chat Conversation']

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?

Answers

The top three categorical/dummy variables in the final model, based on the absolute value of their coefficient factors, are

‘ Total Time Spent on Website, Tags, What is your current occupation. ‘ Total Time Spent on Website, Tags, What is your current occupation ‘, are derived from encoding the original categorical variable ‘Tags’. Similarly, ‘Tags “ , is obtained by encoding the categorical variable ‘Lead Quality’.

Total Time Spent on Website has a coefficient factor of 4.374200 .

What is your current occupation has a coefficient factor of 3.201829 .

Tags has a coefficient factor of 2.573936

- The final model achieves a sensitivity of 0.80, meaning it accurately predicts 80% of the customers who have converted (positive conversion).
- The model has a precision of 0.67, indicating that 67% of the leads predicted as Hot Leads are indeed true Hot Leads.
- we have developed a reusable code block that predicts the conversion value and Lead Score, given the training data, test data, and a specified cutoff. This flexibility allows for the adjustment of cutoffs based on specific use cases, such as when high sensitivity is needed or when optimizing precision is important.

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.

Answers

To make the lead conversion more aggressive during the internship period, X Education should focus on maximizing their outreach to potential leads. Given the goal of converting as many predicted leads (customers predicted as 1) as possible, here is a suggested strategy:

Strategy for Aggressive Lead Conversion

Lower the Cutoff Threshold:

Currently, the model uses a cutoff of 0.1, meaning that any lead with a conversion probability above 10% is predicted as a potential conversion (Convert_predicted = 1).

Lowering the cutoff threshold further (e.g., to 0.05 or 0.03) will increase the number of leads classified as potential conversions. This would identify more prospects to target, aligning with the goal of aggressive outreach.

Prioritize by Lead Score:

Use the Lead_Score to prioritize outreach efforts. Leads with higher Lead_Score values have higher conversion probabilities and should be contacted first.

Sort the leads based on the Lead_Score in descending order and allocate the interns to contact these leads systematically.

Assign Interns Efficiently:

With 10 interns available, distribute the leads evenly based on their scores.

For example, if you have 1000 potential leads, each intern would be responsible for contacting 100 leads. Adjust this number based on the actual lead count and interns' capacity.

Monitor and Adjust:

Continuously monitor the response and conversion rates during this period.

If certain interns are performing better, consider reallocating some leads to them to maximize conversions.

Implementation Example

Using the provided dataset results, the strategy can be implemented as follows:

Lower Cutoff and Predict:

Adjust the model to use a lower cutoff

2. **Sort Leads by Lead Score:** Sort the test data based on the Lead_Score:
- 3 . **Distribute Leads to Interns:** Divide the sorted leads among the interns:

By implementing this strategy, X Education can ensure that their interns focus on the most promising leads, maximizing the chances of converting potential leads during the aggressive conversion period.

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

To minimize the rate of unnecessary phone calls during periods when the company has already met its targets and wants the sales team to focus on other tasks, the strategy should be to set a high cutoff value for predicting lead conversions. This approach ensures that only the most promising leads are

contacted, significantly reducing the likelihood of wasting time on leads that are unlikely to convert.

Strategy to Minimize Unnecessary Phone Calls

Set a High Cutoff Threshold:

Increase the cutoff threshold to a higher value, such as 0.9. This means that only leads with a conversion probability greater than 90% will be classified as likely to convert ($\text{Convert_predicted} = 1$).

This will drastically reduce the number of leads classified as potential conversions, focusing efforts only on the most promising leads.

Model Adjustment:

Adjust the model to use a higher cutoff value for predictions:

python

Copy code

```
build_model_cutoff(X_train[col], y_train, X_test[col], y_test, cutoff=0.9)
```

Focus on High Lead Scores:

Prioritize leads with the highest Lead_Score to ensure that the sales team's efforts are directed towards the most promising prospects.

Sort the leads based on their Lead_Score in descending order and contact only those with the highest scores.

Implementation Example

Here is an example of how you might implement this strategy:

Adjust Cutoff and Predict:

code

```
build_model_cutoff(X_train[col], y_train, X_test[col], y_test, cutoff=0.9)
```

Sort Leads by Lead Score:

code

```
sorted_leads = test_data.sort_values(by='Lead_Score', ascending=False)
```

```
high_priority_leads = sorted_leads[sorted_leads['Convert_predicted'] == 1]
```

Assign Leads to Sales Team:

Provide the sales team with a list of high-priority leads:

code

```
print(high_priority_leads[['Prospect ID', 'Convert_Probability', 'Lead_Score']])
```

Model Evaluation Metrics with High Cutoff

Using a cutoff of 0.9, the model evaluation metrics indicate:

Confusion Matrix:

```
[[1721  13]
```

```
[ 370 619]]
```

```
array([[1340, 371],
```

```
       [ 238, 774]], dtype=int64)
```

Accuracy	61 %
PRECISION	74 %

SENSITIVITY	80 %
SPECIFICITY	72 %

1340: True Negatives (correctly predicted non-conversions).

371: False Positives (incorrectly predicted conversions).

238: False Negatives (incorrectly predicted non-conversions).

774: True Positives (correctly predicted conversions).

Accuracy: 0.776 (77.6%)

Sensitivity: 0.80 (80%) - Lower sensitivity due to the high cutoff, which is acceptable as the focus is on reducing unnecessary calls.

Specificity: 0.72 (72%) - High specificity, indicating a very low false positive rate.

Precision: 0.67598 (67.59%) - High precision, ensuring that most of the predicted conversions are actual conversions.

Conclusion

By implementing this strategy, the company can effectively minimize unnecessary phone calls, allowing the sales team to focus on other important tasks. This approach leverages a high cutoff value to ensure that only the most promising leads are contacted, thereby reducing wasted effort and increasing efficiency.