Lead Scoring Case Study

Subjective Questions

Obtained Logistic Regression Model Coefficients:

Attribute/Column Name	Coefficient	P-Value
'Total Time'	3.324	0.000
'Source_Welingak Website'	5.658	0.000
'Last Activity_Email Bounced'	-1.525	0.002
'Last Activity_SMS Sent'	2.170	0.000
'Current Occupation_Unknown'	-0.858	0.000
'Tags_Closed by Horizzon'	7.098	0.000
'Tags_Lost to EINS'	6.807	0.000
'Tags_Ringing'	-3.596	0.000
'Tags_Will revert after reading the email'	4.611	0.000
'Tags_switched off'	-4.207	0.000
'Last Notable Activity_Modified'	-1.721	0.000

Question 1

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

Answer

The contribution of variables towards prediction of the final probability is dependent on the magnitude of the coefficient of that variable. From the above table summarizing the wrights of the coefficients for each column, we observe that the following variables have the highest absolute value for the coefficient:

- “Tags_Closed by Horizzon” (weight = 7.098)
- “Tags_Lost to EINS” (weight = 6.807)
- “Source_Welingak Website” (weight = 5.658)

Hence these three factors contribute the most towards the probability of a lead getting converted.

Question 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

Since the final model has only one numeric variable and the rest are all categorical variables, we notice that all the three top categorical variables are actually top variables in general. Hence, the three variables from the previous answer (which all happen to be categorical), should be focused on in order to increase the probability of lead conversion. I.e., the variables are:

- “Tags_Closed by Horizzon” (weight = 7.098)
- “Tags_Lost to EINS” (weight = 6.807)
- “Source_Welingak Website” (weight = 5.658)

Question 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

During the period of aggressive pursuance of leads, the sales team can follow some of the following suggestions:

1. The team can focus on leads with lower probability of conversion required for a lead to be classified as "Hot lead" so as to not miss out any leads and achieve maximum conversions.
2. We observe from our obtained model that Tags, Last Activity, Current Occupation, Source of Lead, Time Spent on the website, and Last Notable Activity are important predictors of lead conversion. The sales team can also target leads based on information about these attributes.
3. Lead Source is an important factor for conversion, therefore, the sales team can invest more on positive sources (i.e., sources that are bringing in the most important leads) and try to generate more number of leads from these sources.
4. Last activity (Sent SMS, Email., etc.,) is also an important factor for conversion. The sales team can invest in these modes of communication (based on which communication channel is the best).
5. The least of leads to be followed can be dynamically adjusted. The threshold for considering a high-scoring lead can be updated once all the initial absolutely high priority leads have been contacted.
6. If the lead has additional metadata which has the date/time of last activity of leads, then they can focus on leads with most recent activity first.
7. The team can also improve quality of lead pursuance by providing training to the interns, monitoring their performance, following-up with the leads in a systematic format, and by collaborating with other teams (such as the marketing team).

Note that:

$$Recall = \frac{True\ Positives}{Total\ number\ of\ Actual\ Positives}$$

Since the company's objective mentioned in the question above is to maximize recall, i.e., we want to identify all the positives and penalize the model the most for false negatives, we can therefore build another model to get "Hot leads" with a lower cutoff probability so as to maximize recall so that the sales team can pursue as many of these leads as possible in order to maximize the probability that all the potential convertible leads are pursued.

Question 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

When the company has already reached its target and the sales team is focused on new works, they needn't focus on all leads and instead focus only on leads that surely result in a conversion. There for they can pursue only the leads that are highly likely to convert (as indicated by the lead score). The can increase the cutoff for probability based on which they target a lead or not.
The sales team can follow the following suggestions:

1. Target leads with higher lead scores.
2. Targeting campaigns that can appeal to more number of leads instead of calling one each at a time. Could conduct webinars, workshops, networking events, etc.
3. Introduce self-service options to customers such as comprehensive FAQs, tutorials, online resources, etc., and share them with the leads so that they can educate themselves about the courses and exhibit interest based on which one-one conversations by the sales team can be initiated.

Note that:

$$Precision = \frac{True\ Positives}{Total\ number\ of\ Predicted\ Positives}$$

Increasing precision aligns with the company's objectives during this phase since they want to only contact leads who will convert and have very low tolerance for misclassifying a potential lead who won't convert as a "hot lead". Hence, based on the objective to maximize precision, the Data Science team can build a different model that has a higher cutoff for probability of conversion so as to maximize the Precision.