Lead scoring case study

Submitted by: Rafeek Ponnandy, Priyanka Chatterjee, Radhika Bansal

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

Answer

	coef	std err	z	P>|z|	[0.025	0.975]
const	-2.0578	0.115	-17.933	0	-2.283	-1.833
TotalVisits	3.0001	0.599	5.009	0	1.826	4.174
Total Time Spent on Website	4.562	0.187	24.41	0	4.196	4.928
Page Views Per Visit	-1.5206	0.412	-3.692	0	-2.328	-0.713
Lead Origin_Lead Add Form	3.8964	0.258	15.08	0	3.39	4.403
Do Not Email_Yes	-1.4038	0.214	-6.565	0	-1.823	-0.985
Last Activity_Converted to Lead	-1.102	0.233	-4.725	0	-1.559	-0.645
Last Activity_Email Bounced	-1.0705	0.449	-2.386	0.017	-1.95	-0.191
Last Activity_Had a Phone Conversation	2.1501	0.703	3.059	0.002	0.773	3.528
Last Activity_Olark Chat Conversation	-1.2018	0.188	-6.379	0	-1.571	-0.833
Last Activity_SMS Sent	1.1805	0.086	13.724	0	1.012	1.349
What is your current occupation_Working Professional	2.4917	0.191	13.029	0	2.117	2.866
Last Notable Activity_Unreachable	2.748	0.798	3.443	0.001	1.184	4.312
Source_Olark Chat	1.4853	0.141	10.571	0	1.21	1.761
Source_Welingak Website	2.2903	1.047	2.187	0.029	0.238	4.342

Based on the absolute coefficients and z-values from the regression output, the top three variables contributing most towards the probability of a lead getting converted are:

- **Total Time Spent on Website:** Coefficient = 4.562, z-value = 24.41
- **Lead Origin_Lead Add Form**: Coefficient = 3.8964, z-value = 15.08
- **Total Visits:** Coefficient = 3.0001, z-value = 5.009

These variables have the highest absolute coefficients and z-values, indicating their strong influence on lead conversion.

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

Based on the regression output, the top 3 categorical/dummy variables to focus on for increasing the probability of lead conversion are:

- **Lead Origin_Lead Add Form:** Coefficient of **3.8964**, indicating a strong positive impact and focusing on all leads from **Add Form** may result in higher conversion
- **What is your current occupation_Working Professional:** Coefficient of **2.4917**, showing a significant positive effect, therefore reaching out to the working professionals would increase probability of lead conversion
- **Source_Welingak Website:** Coefficient **2.2903. The leads from source Welingak Website** might have a higher probability of conversion as indicated by the coefficient.

Note: **Last Notable Activity_Unreachable** has higher a Coefficient of 2.748, but its not actionable.

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

- **Focus on High-Probability Leads**: Concentrate on leads with a high probability score (predicted as 1 by the model) as these are the most promising prospects.
- **Create and Prioritize Segments**: Segment the leads based on key variables such as Total ***Time Spent on Website, Lead Origin_Lead Add Form, TotalVisits, What is your current occupation_Working Professional, and Source_Welingak Website***, as these variables have the strongest influence on conversion. Focus on those who spend more time on the website, those from the lead add form and Welingak Website, among others. Each segment can be scored based on their respective values in these variables.
- **Leverage Marketing Strategies**: Utilize various marketing strategies like phone call campaigns, tailored to each lead segment. This includes personal calls, automated calls, SMS, email, etc. Follow up with leads to show the company's seriousness and sense of responsibility towards customers. Customize messaging by sharing success stories, benefits, and testimonials, targeting these messages based on lead segmentation.

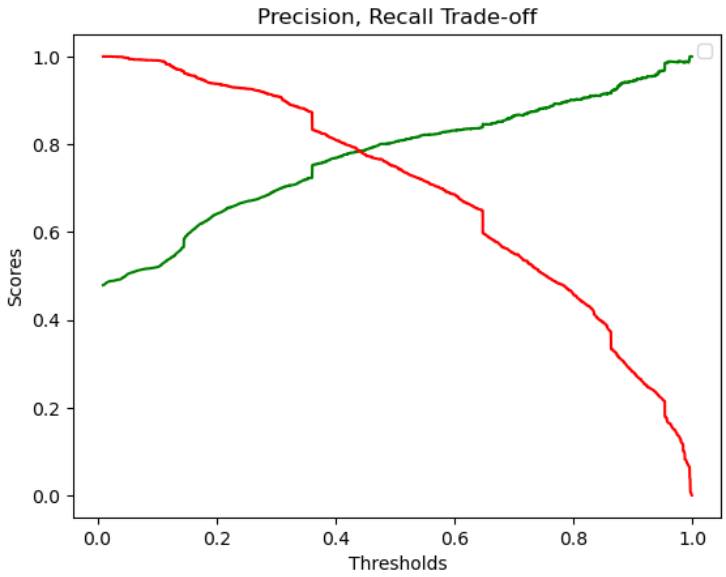
4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

Answer

Focus on Leads Certain for Conversion

From the model output, we should target cut-off probabilities with higher precision. The precision-recall tradeoff chart indicates that at around 80% thresholds, precision is 90% and above. Therefore, we should define the ideal cut-off score for leads to be contacted accordingly.

The chart below shows the model performance at different thresholds.



The below tables show the distribution of model accuracy, sensitivity and specificity at different thresholds based on the training data

	prob	accuracy	sensi	speci
0.0	0.0	0.479083	1.000000	0.000000
0.1	0.1	0.557855	0.991175	0.159334
0.2	0.2	0.718959	0.937761	0.517727
0.3	0.3	0.765243	0.909893	0.632208
0.4	0.4	0.792390	0.810033	0.776164
0.5	0.5	0.792835	0.749652	0.832550
0.6	0.6	0.783044	0.686020	0.872277
0.7	0.7	0.743881	0.551788	0.920547
0.8	0.8	0.716066	0.457966	0.953439
0.9	0.9	0.647530	0.281003	0.984622