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

Answer

I identified 5 variables which contributes most towards the probability.

1. Lead Profile,
2. Lead Quality,
3. Occupation
4. Tags,
5. Last Notable Activity

VIF (Variance Influence Factors) VIF

Encoded Field Name	VIF	Field Name	Meaning of Field
Tags__6	1.03	Tags	Busy
LeadPro__3	1.02	Lead Profile	Lateral Student
Tags__7	1.02	Tags	Switched Off
Tags__4	1.01	Tags	Lost to EINS
Tags__8	1	Tags	In touch with EINS
Tags__25	1	Tags	Want to take admission but has financial problems
LeadQ__5	0.69	Lead Quality	High in Relevance
Occu__3	0.58	Occupation	Working Professional
LeadPro__5	0.4	Lead Profile	Student of SomeSchool
NotableAct__9	0.23	Last Notable Activity	SMS Sent
Tags__3	0.14	Tags	Will revert after reading the email
Tags__9	0.08	Tags	Already a Student
Tags__12	0.02	Tags	Closed by Horizzon

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

	coef	std err	z	P>|z|	[0.025	0.975]
const	-3.1018	0.090	-34.510	0.000	-3.278	-2.926
NotableAct_9	1.2807	0.105	12.242	0.000	1.076	1.486
LeadPro_3	2.9688	1.138	2.609	0.009	0.738	5.199
LeadPro_5	-1.0175	0.328	-3.098	0.002	-1.661	-0.374
LeadQ_5	1.8618	0.237	7.857	0.000	1.397	2.326
Tags_3	3.9230	0.103	38.124	0.000	3.721	4.125
Tags_4	6.1101	0.430	14.208	0.000	5.267	6.953
Tags_6	2.6443	0.206	12.823	0.000	2.240	3.048
Tags_7	-1.9192	0.720	-2.666	0.008	-3.330	-0.509
Tags_8	2.0505	0.700	2.931	0.003	0.679	3.422
Tags_9	-1.2716	0.469	-2.713	0.007	-2.190	-0.353
Tags_12	5.6747	0.401	14.142	0.000	4.888	6.461
Tags_25	2.3179	0.980	2.364	0.018	0.396	4.239
Occu_3	1.8978	0.232	8.196	0.000	1.444	2.352

Top three top dummy variables based on the value of coefficients are

Tags_4 => Lost to EINS
Tags_12=> Closed by Horizzon
Tags3 => Working Professional

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: Our model is predicting 91% accuracy and 92% recall. This is based on the cut off of 65% lead score.
Company X can do the prediction by adjusting this cut-off percentage. Initially they should adjust this cut-off percentage at such a level which gives them the highest accuracy. (On our test data it is 65% right now.) Then call only those leads which are falling above this cut-off range. Once

that list is exhausted then this cut of range should be increased, this will decrease recall and increase specificity. Then you get another list of leads. Excluding those leads which has been contacted earlier filter the difference and call the people from this list.

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:

- Using first list, which comes from adjusting cut-off rate at such a level where we get the highest accuracy, Company X can contact all the hot leads.
- During contacting they should take the feedback, makes notes about the bottlenecks in converting the lead or helping hot leading in making decision.
- After contacting with this list is over sales team should do the followup hot leads using their notes and help them in making decisions.