

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	-0.9661	0.283	-3.417	0.001	-1.520	-0.412
Do Not Email	-1.3127	0.211	-6.223	0.000	-1.726	-0.899
Lead Origin_Lead Add Form	1.0963	0.366	2.995	0.003	0.379	1.814
Lead Source_Welingak Website	3.4147	0.820	4.166	0.000	1.808	5.021
What is your current occupation_Unemployed	-1.1746	0.240	-4.899	0.000	-1.644	-0.705
Tags_Busy	3.8305	0.330	11.598	0.000	3.183	4.478
Tags_Closed by Horizzon	7.9914	0.763	10.480	0.000	6.497	9.486
Tags_Lost to EINS	9.2178	0.755	12.217	0.000	7.739	10.697
Tags_Ringing	-1.6472	0.337	-4.885	0.000	-2.308	-0.986
Tags_Will revert after reading the email	3.9881	0.229	17.380	0.000	3.538	4.438
Tags_switched off	-2.2412	0.587	-3.816	0.000	-3.392	-1.090
Lead Quality_Not Sure	-3.3158	0.129	-25.690	0.000	-3.569	-3.063
Lead Quality_Worst	-3.9600	0.836	-4.734	0.000	-5.599	-2.321
Last Notable Activity_SMS Sent	2.7443	0.120	22.856	0.000	2.509	2.980

the top 3 variables that contribute most towards the probability of a lead getting converted are:

1. Tags\_Lost to EINS
2. Tags\_Closed by Horizzon
3. Tags\_Will revert after reading the email

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: the top 3 categorical/dummy variables that contribute the most towards the probability of a lead getting converted are also :

1. Tags\_Lost to EINS
2. Tags\_Closed by Horizzon
3. Tags\_Will revert after reading the email

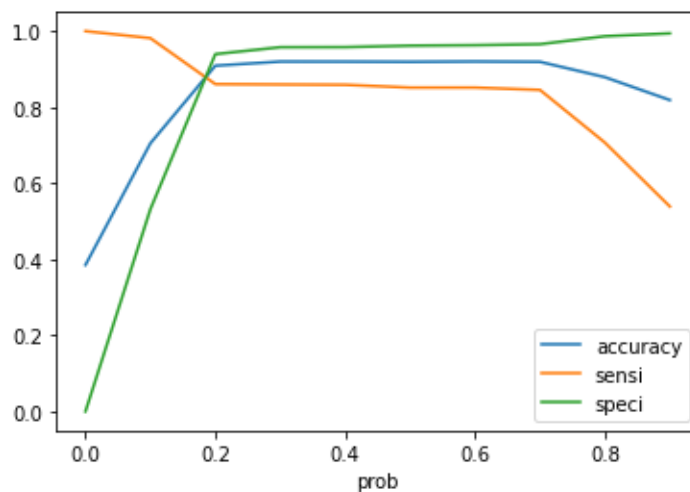
3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers 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: **Sensitivity** with respect to our model can be defined as the ratio of total number of actual Conversions correctly predicted to the total no of actual Conversions.

Similarly, **Specificity** can be defined as the ratio of total no of actual non-Conversions correctly predicted to the total number of actual non-Conversions.

For a particular model, as one increases, the other decreases and vice versa. Different values of the sensitivity and specificity can be achieved for the same model by changing the Conversion Probability cutoff threshold value.

```
# Let's plot accuracy sensitivity and specificity for various probabilities.  
cutoff_df.plot.line(x='prob', y=['accuracy', 'sensi', 'speci'])  
plt.show()
```



When the probability thresholds are very low, the sensitivity is very high and specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.

High sensitivity implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the Conversion likelihood, i.e. it will misclassify some non-Conversion cases as Conversions.

Now, since X Education has more man-power for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all of the potential leads, we can **choose a lower threshold value for Conversion Probability**.

This will ensure the Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible.

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 probability thresholds are very low, the sensitivity is very high and specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.

High sensitivity implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the Conversion likelihood, i.e. it will misclassify some non-Conversion cases as Conversions.

Now, since X Education has more man-power for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all of the potential leads, we can **choose a lower threshold value for Conversion Probability**.

This will ensure the Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible.