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



From the above stats model output we can deduce that the variable **“Total Time Spent on Website”** is the one which leads to increase of the probability of lead conversion.

1. 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?



From the above stats model we can deduce that, dummy variables “**Lead\_score\_\_Live Chat**”, “**Last\_Score\_\_Email Received**” and “**Last\_activity\_\_Resubscribed to emails**” are the ones which will help us achieve higher

1. 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.

|  |  |  |  |
| --- | --- | --- | --- |
| prob | accuracy | sensi | speci |
| 0 | 0.380232 | 1 | 0 |
| 0.1 | 0.56859 | 0.976411 | 0.318388 |
| 0.2 | 0.69293 | 0.881591 | 0.577185 |
| 0.3 | 0.783152 | 0.827012 | 0.756243 |
| 0.4 | 0.794056 | 0.783996 | 0.800227 |
| 0.5 | 0.777348 | 0.628122 | 0.868899 |
| 0.6 | 0.772423 | 0.531915 | 0.919977 |
| 0.7 | 0.764509 | 0.452359 | 0.956016 |
| 0.8 | 0.737249 | 0.350601 | 0.974461 |
| 0.9 | 0.698558 | 0.222942 | 0.990352 |

As during the above-mentioned time, we have interns available who can chase down leads which we might not have considered before, we will set the cut off of the probability to 0.2 or 0.3 which will give the sensitivity of 0.88 and 0.82 respectively.

Below is the modification for the code.

y\_pred\_final['final\_predicted'] = y\_pred\_final.Converted\_Prob.map(lambda x: 1 if x > 0.2 else 0)

OR

y\_pred\_final['final\_predicted'] = y\_pred\_final.Converted\_Prob.map(lambda x: 1 if x > 0.3 else 0)

Output compression:

Output of confusion matrix With cutoff set to 0.4 which is the optimum point

|  |  |  |
| --- | --- | --- |
| #Predicted | not\_converted | converted |
| #Actual |  |  |
| #not\_converted | 1230 | 327 |
| #converted | 174 | 707 |

Output of confusion matrix with cutoff set to 0.2 as per our deduction.

|  |  |  |
| --- | --- | --- |
| #Predicted | not\_converted | converted |
| #Actual |  |  |
| #not\_converted | 876 | 681 |
| #converted | 104 | 777 |

**We see that number of leads is 327+707 ie 1034. But with probability with 0.2 the number of leads is 1458**

1. 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.

|  |  |  |  |
| --- | --- | --- | --- |
| prob | accuracy | sensi | speci |
| 0 | 0.380232 | 1 | 0 |
| 0.1 | 0.56859 | 0.976411 | 0.318388 |
| 0.2 | 0.69293 | 0.881591 | 0.577185 |
| 0.3 | 0.783152 | 0.827012 | 0.756243 |
| 0.4 | 0.794056 | 0.783996 | 0.800227 |
| 0.5 | 0.777348 | 0.628122 | 0.868899 |
| 0.6 | 0.772423 | 0.531915 | 0.919977 |
| 0.7 | 0.764509 | 0.452359 | 0.956016 |
| 0.8 | 0.737249 | 0.350601 | 0.974461 |
| 0.9 | 0.698558 | 0.222942 | 0.990352 |

As during the above-mentioned time, we want to chase only the leads which have highest conversion rate, this can be achieved by going with higher specificity. By looking at above table we should go with 0.8 or 0.9 as the cutoff.

Output compression:

Output of confusion matrix With cutoff set to 0.4 which is the optimum point

|  |  |  |
| --- | --- | --- |
| #Predicted | not\_converted | converted |
| #Actual |  |  |
| #not\_converted | 1230 | 327 |
| #converted | 174 | 707 |

Output of confusion matrix with cutoff set to 0.8 as per our deduction.

|  |  |  |
| --- | --- | --- |
| #Predicted | not\_converted | converted |
| #Actual |  |  |
| #not\_converted | 1505 | 52 |
| #converted | 579 | 302 |

**We see from above with 0.4 probability the number of leads is 1034. But at 0.8 probability the number of leads is 354.**