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

**Ans.:**

Based on the coefficient values obtained from our final model, following are the top three variables that contribute most towards the probability of a lead getting converted :

1. Total Time Spent on Website
2. Lead Origin\_Lead Add Form (dummy variable of original feature: Lead Origin)
3. Occupation\_Working Professional (dummy variable of original feature: What is your current occupation)

Below are the overall features used in the model (sorted by their coefficient values).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **z** | **P>|z|** | **[0.025** | **0.975]** |
| **Total Time Spent on Website** | 4.5655 | 0.161 | 28.419 | 0 | 4.251 | 4.88 |
| **Lead Origin\_Lead Add Form** | 4.1659 | 0.195 | 21.399 | 0 | 3.784 | 4.547 |
| **Occupation\_Working Professional** | 3.8881 | 0.194 | 20.048 | 0 | 3.508 | 4.268 |
| **Occupation\_Other** | 1.7448 | 0.784 | 2.225 | 0.026 | 0.208 | 3.282 |
| **Occupation\_Unemployed** | 1.3252 | 0.083 | 15.935 | 0 | 1.162 | 1.488 |
| **Lead Source\_Olark Chat** | 1.2233 | 0.111 | 10.979 | 0 | 1.005 | 1.442 |
| **Occupation\_Student** | 1.1531 | 0.211 | 5.475 | 0 | 0.74 | 1.566 |
| **TotalVisits** | 1.0453 | 0.233 | 4.489 | 0 | 0.589 | 1.502 |
| **Do Not Email** | -1.3141 | 0.16 | -8.203 | 0 | -1.628 | -1 |
| **const** | -3.295 | 0.113 | -29.048 | 0 | -3.517 | -3.073 |

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?

**Ans.:**

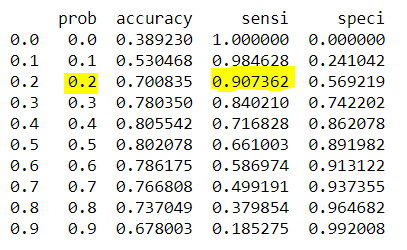
Based on the coefficient values obtained from our final model, following are the top three variables that should be focused most in order to increase the probability of lead conversion:

1. Lead Origin\_Lead Add Form
2. Occupation\_Working Professional
3. Occupation\_Other
4. 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.

**Ans.:**

In order to obtain maximum potential leads the sensitivity of the model can be increased. It will result in capturing a greater number of positive leads. This would increase the false positives (incorrectly identified potential leads) but will reduce the false negatives. Since the company has higher man power during these 2 months they can allow some false positives and have less chances of missing the potential leads. This will improve overall conversion rate.

The model can be tweaked to achieve high sensitivity by reducing the probability cutoff. The cutoff of 0.2 shows good balance of high sensitivity and tolerably low accuracy and specificity.



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.

**Ans.:**

In order to obtain accurate leads the specificity of the model can be increased. It will result in better filtering out non potential leads and reduce false positives (leads incorrectly identified as potential). This will increase the likelihood of a customer being a potential lead and minimize the rare of useless phone calls.

The model can be tweaked to achieve high specificity by increasing the probability cutoff. The cutoff of 0.6 shows good balance of high specificity and accuracy along with tolerably low sensitivity.

