Logistic Regression Assignment

Subjective Questions

By – Kaivalya Degaonkar

Khushi Somaiya

Kritik Sahu

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

   The top variables which contribute most towards the probability of a lead getting converted can be found out by looking at the coefficients of the variable in the final model from the summary.

   The top three variables in the model which contribute most towards the probability of a lead getting converted are:
   - Lead Origin_Lead Add Form
   - Lead Source_Welingak Website
   - What is your current occupation_Working Professional

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?

   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 are:
   - Lead Origin_Lead Add Form
   - Lead Source_Welingak Website
   - What is your current occupation_Working Professional

   The answers to both the above questions are same since in our final model the top 3 variables are all categorical/dummy variables.

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.

   Sensitivity is number of actual Yeses correctly predicted out of total number of Yeses.

   By changing the values of cut-off threshold for probability of lead conversion, the values of sensitivity obtained can also be different. Based on this we can develop a good strategy.

   X Education has more man power for 2 months so, they can get in contact with more customers than before to try increasing the conversion rate. To increase the conversion rate which in turn means to increase sensitivity, we can lower the threshold value in our model. Right now for threshold value of 0.37, the

sensitivity is about 81%. For example, to increase the sensitivity to 86% we can choose a lower value such as 0.3 as the threshold value. The below table can be referred to know the values of sensitivity based on different threshold(probability) values for our model.

Probability	Accuracy	Sensitivity	Specificity
0.0	0.381487	1.000000	0.000000
0.1	0.607437	0.972625	0.382195
0.2	0.766614	0.910825	0.677667
0.3	0.806646	0.867275	0.769250
0.4	0.815823	0.785566	0.834485
0.5	0.813449	0.698051	0.884625
0.6	0.806487	0.625467	0.918138
0.7	0.783861	0.515968	0.949092
0.8	0.761234	0.422231	0.970325
0.9	0.717089	0.279137	0.987209

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.

   Specificity means number of actual Nos correctly predicted out of total number of Nos. Based on this, we can develop a good strategy for the given situation. It gives us idea about non-conversion rate.

   Since the company has already achieved its target for a quarter before the deadline and wants to avoid unnecessary calls, we can try increasing the value of specificity.

   Increasing specificity may increase the chances of misinterpretation of possible conversions as non-conversions but this way the sales team can focus on only extremely necessary phone calls. The sales team can use the saved time to work for their new tasks.

   To increase the value of specificity, we can try increasing the threshold value. The above table can be referred to check for values of different specificity based on different threshold(probability) values for our model.