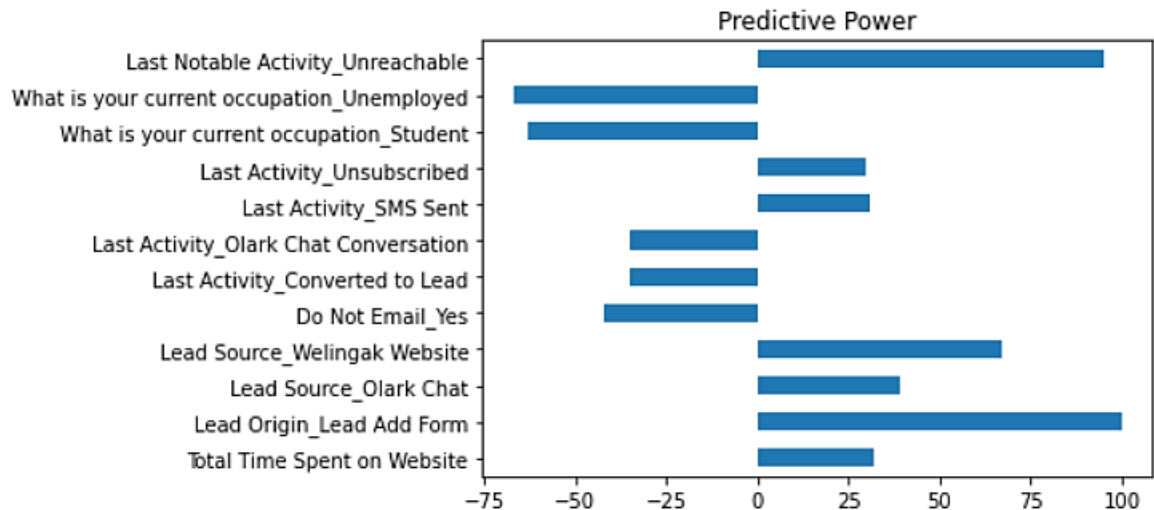


Lead Scoring for X Education

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



From the graph, we could see that the top 3 variables that contribute most towards the probability of lead conversion are:

- 1 → Lead Origin_Lead Add Form
- 2 → Last Notable Activity_Unreachable
- 3 → Lead Source_Welingak Website

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?

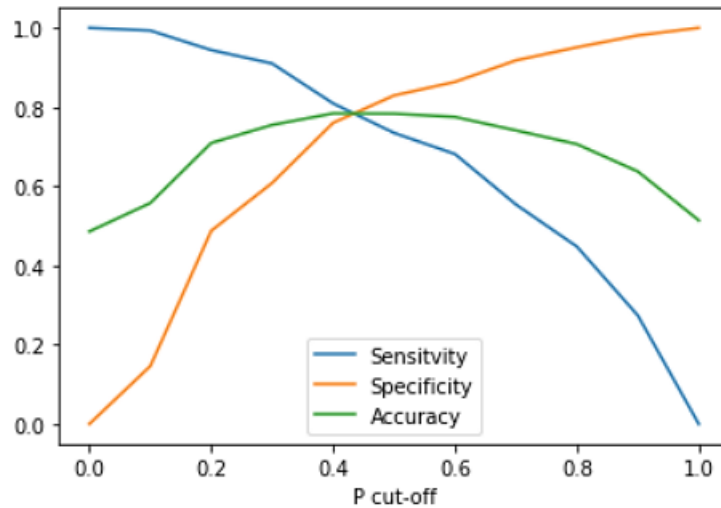
The top 3 categorical/dummy variables that are to be focused the most in order to increase the probability of lead conversion are same as the top 3 variables that contribute the most. This is because we have only categorical variables in the top 3 and, no numerical variables.

So, here are those top 3 categorical/dummy variables:

- 1 → Lead Origin_Lead Add Form
- 2 → Last Notable Activity_Unreachable
- 3 → Lead Source_Welingak Website

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 focuses on measuring the “true” or “1” values correctly. Specificity focuses on measuring the “false” or “0” values correctly. The correctness of sensitivity & specificity is decided based on the precision - recall trade-off, by calculating the F1 score.



Sensitivity decreases with increase in probability. Specificity increases with increase in probability.

We now have sufficient man-hours, allowing more room for failure in prediction, so we can increase the sensitivity of our model. Considering the sensitivity vs. probability plot as an elbow-curve, we could see that there is a shift in its linearity at P cut-off = 0.2 and P cut-off = 0.1 after P cut-off = optimal. We may choose either of these two points as our new cut-off (lower) for probability, to have high sensitivity.

In this scenario, we will consider almost all leads for conversion, including those having marginal likelihood, due to sufficient man-hours.

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.**

We now have less man-hours, allowing less room for failure in prediction, so we can increase the specificity of our model. Considering the specificity vs. probability plot as an elbow-curve, we could see that there is a shift in its linearity at P cut-off = 0.5, P cut-off = ~0.6 and P cut-off = ~0.7 after P cut-off = optimal. We may choose either of these three points (0.5 is preferred, since other two are approximate) as our new cut-off (higher) for probability, to have high specificity.

In this scenario, we will identify only those leads having high likelihood for conversion, leaving behind the ones with moderate to low likelihood, due to less man-hours.