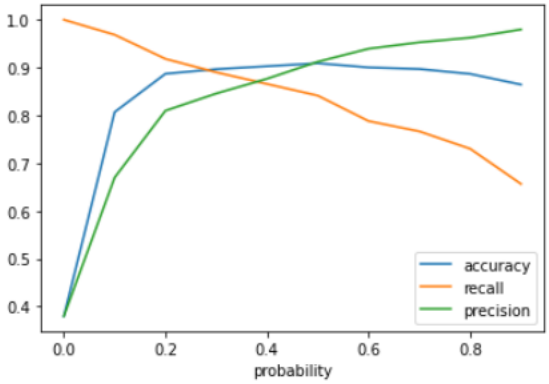
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?
    1. *Tags*
    2. *Lead Origin*
    3. *Total Time Spent on Website*

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?
    1. *Tags_Closed by Horizzon*
    2. *Tags_Will revert after reading the email*
    3. *Lead Origin_Lead Add Form*

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.
    To ensure that the lead conversion is more aggressive, the focus would be to ensure that all the hot leads are converted. This can be achieved if we consider a lower cutoff of the lead score so that more leads can be considered for calling due to the available bandwidth. This would help to cover as many hot leads and increase the True Positives. This will lead to higher accuracy. However, this may cause slightly lower recall and precision values. The preferred cutoff from the metric graph for this case would be between 0.4 to 0.5 for probability (Lead scores of 40 to 50).



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.
    Since the sales bandwidth is limited, we have to ensure that high level of lead conversion is achieved by converting as many of the hot leads. However, due to the limited sales bandwidth,

the calls to cold leads have also to be reduced. This means that the true positives have to be increased while reducing the false positives and false negatives. This is achieved by increasing the precision and recall values. So, based on the above graph, we can consider cutoff value between 0.25 to 0.4 for probability (Lead scores of 25 to 40).