# Subjective Questions

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

Ans: The 3 key variables according to my model are:

1. Tags\_Closed by Horrizon
2. Tags\_Lost to EINS
3. Lead Source\_ Welingak Website

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: The 3 key variables according to my model are:

1. Tags\_Closed by Horrizon
2. Tags\_Lost to EINS
3. Lead Source\_ Welingak Website

Assuming we know the meanings of these Tags (not given in the data dictionary) we will target this. Otherwise we will target:

1. Lead Source\_Welingak Website – pursue the people who come from here
2. Last Activity\_SMS Ent – pursue the people whose last activity is SMS sent
3. Lead Quality\_Worst – if employees intuition suggests candidate isn’t interested, then don’t pursue them at all
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.**

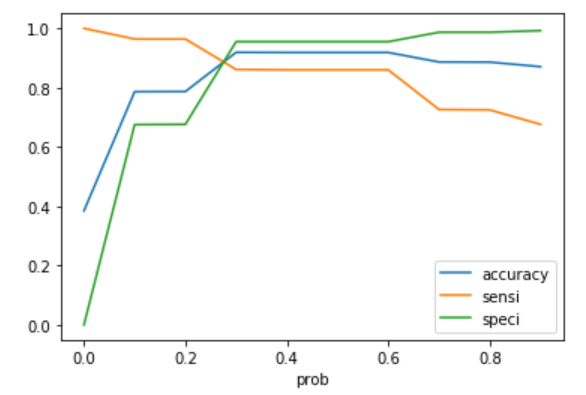
Ans:

For answering, the above scenario we need to understand what is sensitivity and specificity in this case of the model.

**Sensitivity**: With respect to our model, it is the ratio of total number of conversions correctly predicted to the total number of actual conversions.

**Specificity**: With respect to our model, it is the ratio of the total number of non-conversions predicted to the total number of actual non-conversions.

Let’s look at the graph for sensitivity vs specificity for our model



Therefore, from the graph above we can infer that at lower cutoff, the sensitivity is high but the specificity is low and at higher cutoff, it is vice versa. High sensitivity implies that model will correctly predict all the leads who are likely to predict though misclassifying some non-conversions as conversion.

Now since XE education has more manpower in the two months they can choose a lower threshold value if they wish to make the lead conversion more aggressive as all the leads who are likely to convert will be identified correctly and they can make phone call to all of them.

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:

We can follow the similar approach as above but in the opposite manner. In the previous question, we have taken lower threshold value for higher sensitivity so that all the leads who are likely to be converted are identified correctly but here we will take higher threshold to get higher specificity so that all the leads who are not likely to be converted are correctly identified.

Therefore, as XE Education has reached its target and does not want to make unnecessary phone calls then we can choose higher threshold value for conversion probability, which will ensure specificity is high, and hence the model will identify all the leads who less likely to be converted identified correctly. As a result, agents will not make unnecessary calls.