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

**Ans-** Top 3 variables in in my model are:

1) Lead Source\_Wellingak Website

2) Lead Source\_Reference

3) Do Not Email

These variables have the highest absolute value among all the coefficients and therefore can be considered as the ‘strongest’ indicators of conversion. The meaning we can derive from this is that if the lead source originates from Wellngak website or from a reference they are most likely to convert. Since ‘Do Not Email’ has a high negative coefficient people who have NOT checked this option would be more open to conversion. This variable has an inverse correlation with the conversion rate.

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 top 3 categorical/dummy variables which should be focused on are:

1) Lead Source\_Wellingak Website

2) Lead Source\_Reference

3) Lead Activity\_SMS Sent

These variables have coefficients with the highest positive value among the dummy variables. What this means is that if the lead source came from the Wellingak website or from a reference they are most likely to be converted. Also we can note that if the person has sent an SMS they are more likely to convert.

Table

Description automatically generated

1. X Education has a period of 2 months every year during which they hire some interns. The sales team 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 many of such people as possible. Suggest a good strategy they should employ at this stage.

**Ans-** So during this time since there is no constraint on the number for phone calls what needs to be focused on is targeting the highest number of potential leads identified by the model. Therefore, we need to ensure that customers who can be converted are not incorrectly classified as non-leads (false negatives). So, we need to choose a threshold value which will increase the recall/sensitivity score. From the notebook we can see that it is possible to achieve this by choosing the threshold as 0.2 or lower which would give a high recall score. Taking it as 0.2 nets a recall score of 88%. This can ensure that most of the potential leads identified by the model are contacted thus improving the conversion rate.

Recall/Sensitivity = True Positives/ True Positives +False Negatives

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 not to 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 this scenario we take the opposite approach from the previous one. We are aiming to reduce the number of false positives, meaning customers who would not have converted incorrectly classified as potential leads. Contacting these customers would waste time since they wouldn’t have converted anyways. Therefore, we wish to increase our precision score.

We can achieve this by choosing a higher threshold value as seen from the precision recall curve. Any value above 0.8 would be suitable in identifying only those leads which will most likely be converted and the number of positive identifications will be less leading to lesser number of overall phone calls but near perfect conversion rate within the given leads approached.

Chart, line chart

Description automatically generated