Subjective Questions with Answers

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

Answer : The top three variables in our model which contributed most towards the probability of lead a getting converted are 1. Total Time Spent on Website 2. Lead Origin_Lead Add Form (Dummy Variable) 3. What is your current occupation_Working Professional (Dummy Variable).

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?

Answer : Top 3 categorical variables are --
a) Lead Origin_Lead Add Form (Dummy Variable)
b) What is your current occupation_Working Professional (Dummy Variable)
c) Lead Source_Welingak Website (Dummy Variable)

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.

Answer : In this scenario, the goal is to maximize the number of leads predicted for conversion by lowering the model's probability cutoff. While this means that some non-potential leads may also be included, it increases recall ensuring that more actual potential leads are identified, even at the cost of precision. This trade-off allows for capturing a larger pool of possible customers, prioritizing lead generation over absolute accuracy.
The confusion matrix helps explain how the model classifies leads:
    True Positive (TP): Correctly predicting a potential lead as a lead.
    False Positive (FP): Incorrectly predicting a non-lead as a lead.
    False Negative (FN): Missing a potential lead by predicting it as a non-lead.
    True Negative (TN): Correctly predicting a non-lead as a non-lead.
Recall measures how well the model identifies actual leads by calculating the proportion of correctly predicted leads out of all actual leads. Precision, on the other hand, evaluates how many of the predicted leads are truly potential customers. By focusing on recall, the strategy ensures that fewer potential leads are overlooked, even if it means a slight drop in precision.

To improve lead conversion, the company should target users who: a)Spend more time on the website b)Fill out the Lead Add Form, indicating strong intent c)Arrive at the site via the Welingak Website, a key lead source d)Are working professionals, as they are more likely to enroll.e)Have a higher number of total visits, showing continued interest. By prioritizing these factors, X Education can enhance its lead generation strategy, ensuring that a larger share of potential customers are engaged and nurtured toward conversion.

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.

Answer : In this approach, the model uses a higher probability cutoff to predict only the leads that are most likely to convert. This increases precision, ensuring fewer false leads, but it also reduces recall, meaning some potential leads might be missed. The focus shifts from quantity to quality, prioritizing accuracy over capturing a larger pool of leads. With this strategy, X Education should concentrate on customers who spend more time on the website, as their higher engagement suggests a stronger intent to enroll.