

Subjective Questions Answer – Lead Conversion Project

By Umar Farooq Bhat

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

Based on both logistic regression coefficients and Random Forest feature importance, the top three predictive variables in the model are:

- 1. Tags_Interested in Next Batch – Strongly associated with high intent to convert.**
- 2. Total Time Spent on Website – Indicates strong engagement; more time often leads to conversion.**
- 3. Lead Source_Olark Chat – Leads from live chat convert at significantly higher rates.**

These features are highly influential and actionable for the business.

2. What are the top 3 categorical/dummy variables in the model which should be focused the most on to increase the probability of lead conversion?

The most important categorical variables in dummy form are:

- Tags_Interested in Next Batch – Immediate follow-up can increase conversion.**
- Tags_Will Revert After Reading Email – Indicates intent but requires nurturing.**
- Lead Source_Olark Chat – Leads from this source are already engaged and closer to conversion.**

Focusing on leads that fall into these categories can directly improve conversion performance.

3. 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 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.

During this aggressive phase, the goal is maximum outreach. The recommended strategy:

- **Lower the prediction threshold (e.g., from 0.5 to 0.3 or 0.4) to capture more potential converters.**
- **Focus on increasing recall — ensure no potential converters are missed.**
- **Interns should call all leads predicted as 1, even those with moderate probability.**
- **Prioritize by predicted probability, so interns call the highest-scoring leads first.**
- **Track results and feedback weekly to refine outreach efforts.**

This approach ensures aggressive but structured conversion, fully utilizing intern capacity.

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.

Here, the focus is efficiency over volume. The recommended strategy:

- **Raise the prediction threshold (e.g., from 0.5 to 0.7 or 0.8) to filter for only high-confidence leads.**
- **This increases precision, reducing calls to low-potential leads.**
- **Use email automation for low-score leads to stay engaged without sales effort.**

- **Sales reps can focus on strategic work: training, data feedback, or CRM updates.**
- **Leads with borderline scores can be tagged for future follow-up**