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

Answer: - Last Notable Activity Had a Phone Conversation 23.436868

Tags Closed by Horizon 7.431015

Tags Lost to EINS 7.161059

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?

Answer: - Lead Source_Welingak Website 2.33

Last Activity Email Opened 1.78

What is your current occupation Working Professional 1.62

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.

Answer: - To make the most of the **intern-heavy 2-month phase** and maximize **lead conversion**, X Education should **adjust their lead scoring and calling strategy** based on model probabilities. Here's a solid plan:

✅ **Strategy for Aggressive Lead Conversion (Intern Period)**

🎯 **Objective:**

Convert **as many high probability leads as possible** by making calls to all potential leads (predicted as likely to convert).

🔧 **1. Lower the probability threshold**

- **Default threshold**: Most logistic models use **0.5** to classify leads as 1 (converted) or 0 (not converted).
- **New strategy**: **Lower the threshold** to **0.3 or 0.25** during this period.

📌 **Why?**

- It **increases recall** (true positives) at the cost of precision (some false positives).
- That's okay during intern phase, because **calling extra people isn't costly** due to intern support.

Example of changing threshold:

y_pred_prob = model.predict(X_test) # model could be res or logistic_model

y_pred_adjusted = (y_pred_proba >= 0.3).astype(int) # use 0.3 threshold

---

📞 2. Prioritize based on probability scores

Even among predicted 1s, **rank leads** by their probability scores:

- Call leads with **higher probabilities first**, then move down the list.
- Create a **call priority queue** for interns.

Assuming you have a DataFrame of predictions

lead_scores['priority'] = y_pred_proba

lead_scores_sorted = lead_scores.sort_values(by='priority', ascending=False)

📈 3. Monitor precision and conversion rates weekly

- Track how many leads are actually converting under the adjusted threshold.
- If too many calls go to unqualified leads, you may need to **adjust the threshold slightly back up** (e.g., from 0.3 to 0.35).

📊 4. Use key variables to filter further

Only select predicted 1s who also meet strong criteria, such as:

- Source = 'Welingak Website'
- Last Activity = 'Email Opened'
- Occupation = 'Working Professional'

These are the **top dummy variables** you identified earlier.

✅ Final Recommendation

Use a lower threshold (like 0.3) to widen the net of predicted conversions. Then, rank leads by their probability scores and prioritize calls. Filter further using top predictors like lead source and last activity. This strategy **maximizes the chances of conversions** without increasing cost, making the most of the intern availability.

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: - When X Education **reaches its sales target early**, the focus shifts from maximizing conversions to **minimizing unnecessary effort** (i.e., making **only essential phone calls**). The goal is to **increase precision**, even if it means missing out on some potential leads.

✅ **Strategy for Minimizing Useless Phone Calls (Post-Target Phase)**

🎯 **Objective:**

Avoid wasting time and resources by calling only the most likely-to-convert leads.

🔧 **1. Increase the probability threshold**

- **Default threshold**: 0.5
- **New strategy**: **Raise the threshold to 0.7 or 0.75**

📌 **Why?**

- This will **reduce false positives** (leads predicted as likely to convert but don't).
- You'll be contacting only the leads with a **very high probability of conversion**.

python

Example: Only classify as "convert" if probability ≥ 0.75

y_pred_conservative = (y_pred_proba >= 0.75).astype(int)

📉 **2. Focus on precision over recall**

Use metrics to **track precision**, not just accuracy:

python

from sklearn.metrics import precision_score

precision_score(y_test, y_pred_conservative)

You want **high precision**, even if **recall drops**. That means most of your calls are productive.

🔍 **3. Filter using top predictors**

Even among high-probability leads, **further narrow down** using variables with high impact:

- **Lead Source**: Prioritize “Welingak Website”
- **Last Activity**: "Email Opened", "SMS Sent"
- **Occupation**: “Working Professional”

These variables are **strong indicators of conversion** and help reduce wasteful calls.

🧠 4. Segment leads for optional calling

- Leads with probability **between 0.5 and 0.7** can be labelled as “Optional”.
- Keep them in CRM but don’t contact unless needed.

✅ Final Recommendation

Raise the probability threshold (e.g., to 0.75) to classify a lead as worth calling. Prioritize precision. Use strong categorical predictors to filter further and optionally ignore borderline leads.

This ensures that **only the most qualified leads are contacted**, helping the team shift to other work without losing core productivity.