

Lead Scoring Case Study

By - Shaik Mahammad Ghouse

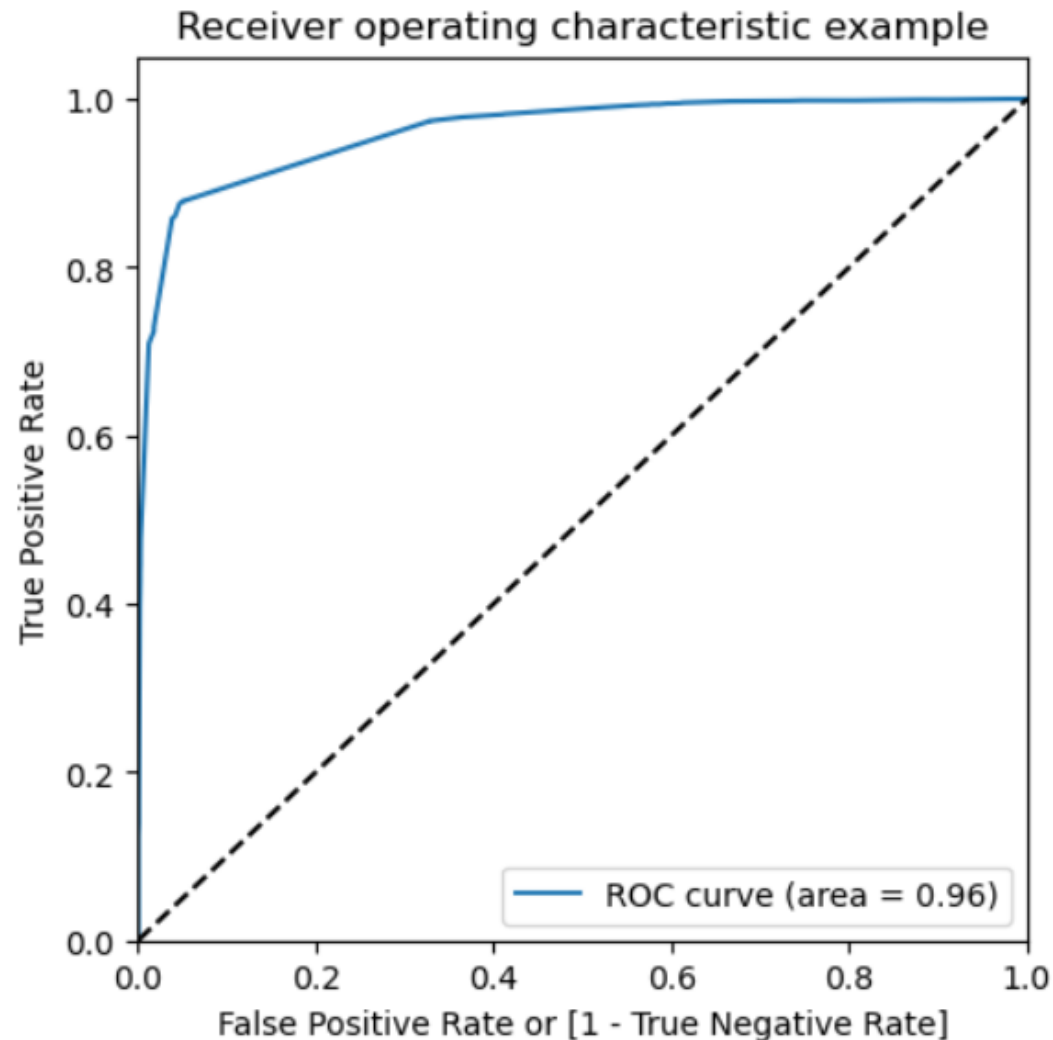
Problem Statement

- X Education wants to increase its lead-to-sale conversion rate by building a logistic regression model that assigns a lead score to each lead. This score will help the sales team identify and focus on high-potential leads, aiming to improve the conversion rate from ~30% to ~80%.

Execution Plan

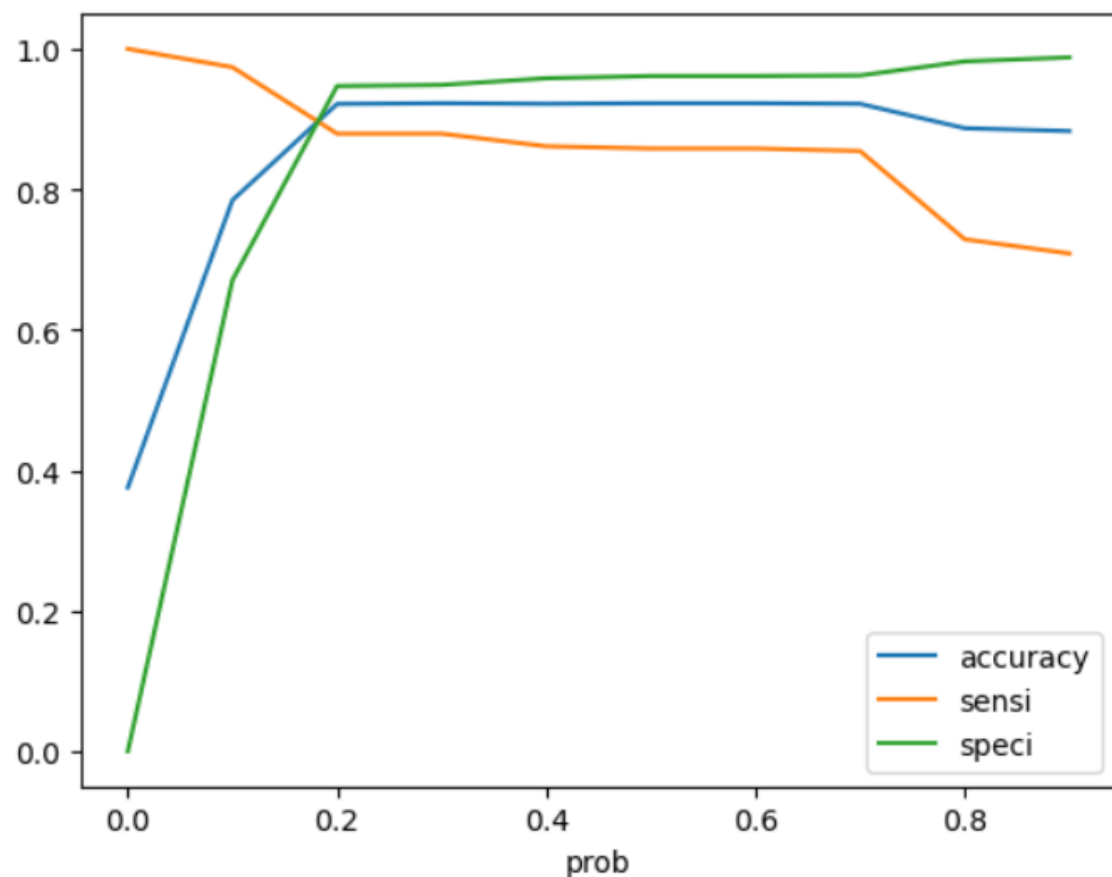
- ❑ Split the cleaned dataset into training and testing sets after addressing missing values.
- ❑ Used **Recursive Feature Elimination (RFE)** to select the top 15 most impactful features after handling dummy variables and scaling.
- ❑ Developed a **Logistic Regression model**, achieving over **92% accuracy** and **87% sensitivity** on the training set.
- ❑ Determined the optimal classification threshold by analyzing accuracy, sensitivity, and specificity trade-offs.
- ❑ Evaluated the model on the test set, maintaining **92% accuracy** and **87% sensitivity**.
- ❑ Assigned a **lead score** to each lead by multiplying the predicted probability by 100, enabling ranking based on conversion likelihood.

ROC Curve



- ❑ The ROC curve indicates that the model achieves an AUC (Area Under Curve) of 96%, reflecting strong predictive performance.
- ❑ The model demonstrates a high accuracy in distinguishing between leads that convert and those that don't.

Optimal Probability Cutoff



- ❑ An optimal threshold value of 0.2 was chosen to improve the model's performance in classifying lead conversions
- ❑ At this cutoff, the model achieves:
 - **Accuracy: 92%**
 - **Sensitivity: 87%**
 - **Specificity: 94%**

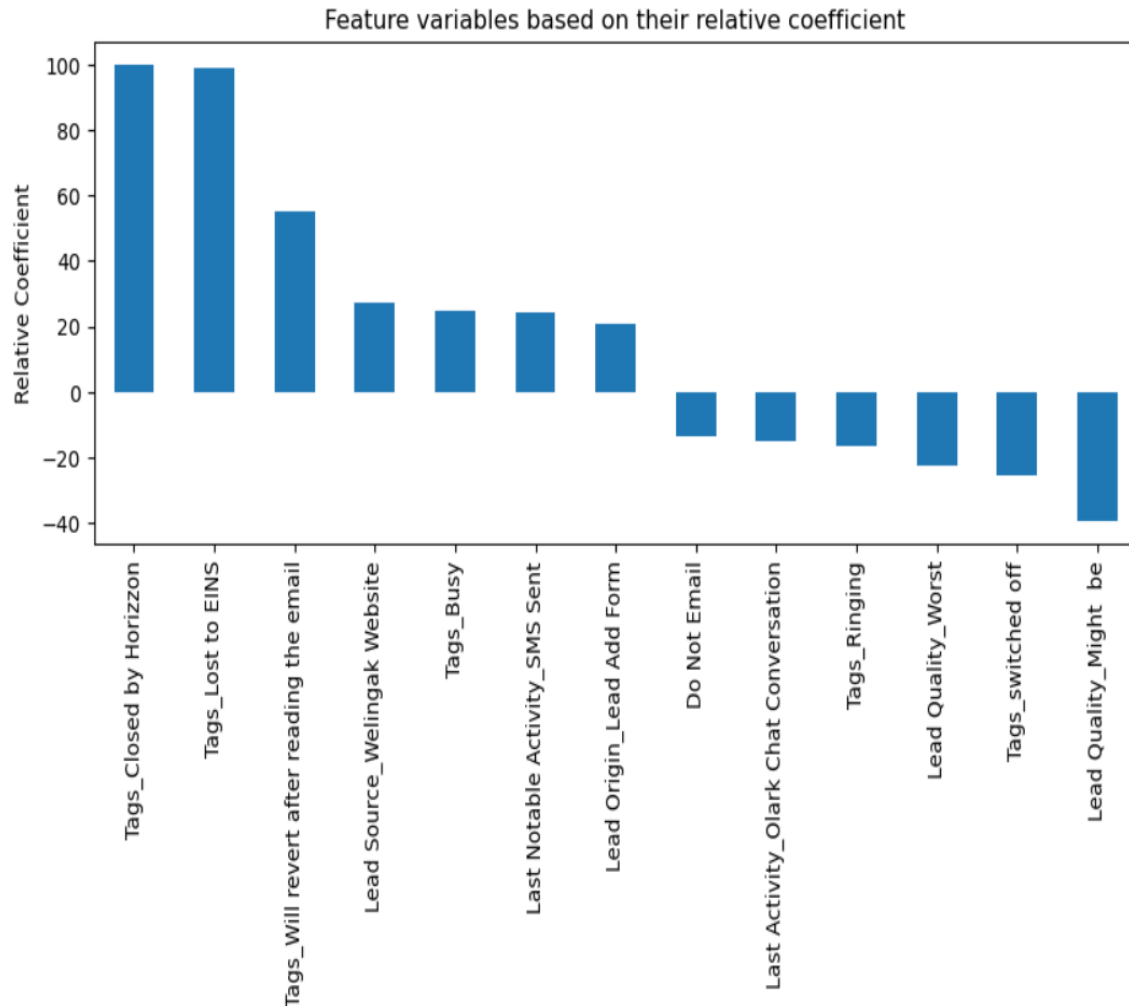
Confusion matrix on Test data

Actual/Predicted	Not Converted	Converted
Not Converted	1498	82
Converted	121	842

Accuracy : 92% | Sensitivity : 87% | Specificity : 94%

- The model demonstrates **92% accuracy** in predicting lead conversions on new, unseen data, enabling the company to reliably identify potential ‘hot’ leads.
- It correctly identifies **87% of the leads that actually convert**, ensuring strong recall of valuable prospects.
- Additionally, the model accurately predicts **94% of non-converting leads**, helping X Education significantly reduce time and effort spent on low-potential leads—ultimately improving overall efficiency and resource allocation.

Important Features



❑ Top 3 Factors Positively Impacting Lead Conversion:

- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Tag_Will revert after reading the email

❑ Top 3 Factors Hindering Lead Conversion:

- Lead Quality_Might be
- Tags_switched off
- Lead Quality_Worst

Suggestion

- ❑ Leads with high scores should be prioritized as "hot" leads. The sales team is encouraged to follow up promptly, as these leads have a high potential for conversion.
- ❑ Leads marked with 'Do Not Email' should be excluded from future outreach, as they've opted out of communication.
- ❑ Leads previously categorized as 'Might Be' or 'Worst' based on chat history can be deprioritized or ignored, given their low conversion likelihood.

Thank You