

LEAD SCORING CASE STUDY SUMMARY REPORT

Problem Statement & Objective

X Education, an online learning company, receives a high volume of leads from various sources. However, not all leads convert into paying customers. The company wanted to develop a lead scoring model that would help identify the most promising leads, enabling the sales team to prioritize their efforts efficiently. The primary goals of this project were:

- To analyze historical data and understand key factors influencing lead conversion.
- To build a logistic regression model that predicts the likelihood of a lead converting into a customer.
- To provide actionable insights to the business team for improving conversion rates and optimizing marketing strategies.

Data Understanding and Preprocessing

The dataset was loaded and examined to understand its structure. The initial steps involved:

- Checking for missing values and deciding how to handle them. Columns with a very high percentage of missing values were removed, while others were imputed based on logical assumptions or statistical measures.
- Identifying and converting categorical variables into numerical representations using one-hot encoding.
- Dropping redundant or non-informative columns such as "Prospect ID" and "Lead Number" since they did not contribute to the prediction process.
- Performing exploratory data analysis (EDA), including visualizing distributions, analyzing correlations between variables, and understanding the impact of different features on lead conversion.

Feature Selection and Model Building

To create an efficient and interpretable model, we used Recursive Feature Elimination (RFE) to select the most important predictors. Logistic regression was chosen for classification due to its simplicity and ease of interpretation. The model was trained using the selected features and evaluated based on key performance metrics:

- Accuracy – Measures the overall correctness of the model.
- Precision & Recall – Important to balance false positives and false negatives.
- F1-score – A harmonic mean of precision and recall to evaluate the overall performance.
- ROC-AUC Score – Used to measure how well the model differentiates between converted and non-converted leads.

Key Findings from Model Insights

The model identified the top factors influencing lead conversion. The three most important features contributing to a higher probability of conversion were:

1. Tags_Lost to EINS – If a lead was previously marked as lost to a competitor (EINS), it strongly affected their probability of converting.
2. Tags_Closed by Horizzon – Leads that were previously closed were less likely to convert.
3. Tags_Will revert after reading the email – Engagement through emails significantly increased the chances of conversion.

Additionally, categorical features like Lead Source, Lead Profile, and Last Activity played a crucial role in determining conversion probability. Leads from certain sources had a much higher likelihood of converting than others.

Business Recommendations & Strategy

1. During Intern Hiring Period (Aggressive Conversion Phase)

Since X Education hires interns for two months every year to assist the sales team, a more aggressive lead conversion strategy should be implemented during this period:

- Prioritize outreach to leads with the highest probability of conversion based on the model's predictions.
- Increase the number of follow-ups, especially through emails and SMS, as these have been found to impact conversions significantly.
- Utilize automated workflows to engage with leads who have shown prior interest, ensuring no potential customer is left unattended.

2. During Target Achievement Period (Reducing Efforts)

When the company has already met its sales targets for the quarter, the goal should be to optimize resources and minimize unnecessary follow-ups:

- Reduce sales efforts on leads that have a low probability of conversion, as determined by the model.
- Focus only on high-quality leads that exhibit strong engagement signals.
- Adjust the lead scoring threshold to ensure that only the most promising leads receive attention, thereby reducing unnecessary phone calls and emails.