

Lead Score Case Study Summary

X Education: How to Increase Conversion Rate of Industry Professionals

Introduction

This document summarizes the lead score case study report that was done for X Education, an online education platform that offers courses to industry professionals. The main objective of the analysis was to find ways to improve the conversion rate of leads, i.e., the percentage of leads who enroll in a course after showing interest. The report followed a data-driven approach to identify the most important factors that influence the conversion rate and to suggest actionable recommendations for X Education.

Methodology

The report used the following steps to conduct the analysis:

- **Cleaning Data:** The data was partially clean except for a few null values and irrelevant categories. The null values were replaced with appropriate values based on the context and the categories were simplified to reduce noise.
- **EDA:** An exploratory data analysis was done to check the distribution and relationship of the variables. It was found that a lot of elements in the categorical variables were inapplicable or had low frequency.
- **Data Preparation:** The data was split into train and test sets, where 70% of the data was used for training and 30% for testing. The categorical variables were converted into dummy variables for modeling.
- **Model Building:** A logistic regression model was built using recursive feature elimination (RFE) with 15 variables. The variables were further selected based on the variance inflation factor (VIF) and the p-value. The final model had 10 variables with VIF less than 5 and p-value less than 0.05.
- **Model Evaluation:** A confusion matrix was made to evaluate the model performance on the train set. The optimal cut-off value was found to be 0.2 by plotting the accuracy, specificity and sensitivity for various probabilities.
- **Prediction:** The model was applied on the test set with the optimal cut-off value of 0.2. The accuracy, sensitivity and specificity of the model on the test set were calculated and compared with the train set.
- **Precision-Recall:** The precision and recall of the model on the test set were calculated and the trade-offs between them were discussed. The precision-recall curve and the area under the curve (AUC) were also plotted to visualize the model performance.

Results

The main results of the analysis are as follows:

- The final logistic regression model had an accuracy of 79.6% on the train set and 79.1% on the test set, indicating a good fit and generalization.
- The model had a sensitivity of 79.5% on the train set and 78.7% on the test set, meaning that it correctly identified 79.5% and 78.7% of the leads who converted, respectively.
- The model had a specificity of 79.7% on the train set and 79.5% on the test set, meaning that it correctly rejected 79.7% and 79.5% of the leads who did not convert, respectively.

- The model had a precision of 76.4% on the test set, meaning that 76.4% of the leads predicted to convert actually converted.
- The model had a recall of 78.7% on the test set, meaning that 78.7% of the leads who actually converted were predicted to convert.
- The model had an AUC of 0.86 on the test set, indicating a high degree of discrimination between the positive and negative classes.
- The top three features for a good conversion rate were: last notable activity had a phone conversation, lead origin lead add form, and what is your current occupation working professional.

Recommendations

Based on the results, the report suggested the following recommendations for X Education:

- Focus on the leads who had a phone conversation as their last notable activity, as they have the highest probability of conversion.
- Invest more resources on the lead add form, as it generates the most qualified leads with a high conversion rate.
- Target the working professionals, as they are the most likely to enroll in a course for upskilling or career advancement.