Lead Score Case Study Report

Problem statement:

The case study describes the steps in developing the lead scoring model. The lead score can be calculated by taking into account various factors such as the lead's likelihood of converting. X Education is a seller of online courses that are designed for industry professionals. It needs help identifying the most promising leads.

Business Goal:

A lead scoring model should be used by the company to assign a lead score to each prospect. It will help determine which leads have a higher chance of converting and which ones have a lower chance. The CEO of the company had stated that the company's objective is to have a lead conversion ratio of 80%.

Solution Summary:

A. The first step in the process is to thoroughly understand the data.

B. The second step involves cleaning the data. To clean the data, we first removed the variables with unique values. Then, we changed the values in the columns with the value "Select" to null. These values indicate that the leads did not choose to follow the given option.

C. We then removed the redundant and imbalanced variables. We also performed a cleanup step by imputing the missing values in the numerical and categorical variables. Outliers were also identified and removed.

D. This issue was solved by changing the label from small to upper case. The sales team's generated variables were also removed from the final solution to avoid any confusion.

E. We then created dummy categorical variables. We also removed the redundant and repeated variables.

F. The next step is to split the data into test and training sections. The proportion of values that will be used for the training sections will be around 70% to 30%.

G. Feature Rescaling. To scale the original variables, we used the Min Max scaling method. We then created a Heatmap to analyze the correlations between the various variables.

H. After selecting the 15 most crucial features, we proceeded with recursively looking at the data to identify the most significant ones. We then came across the 11 most vital variables. The results of the analysis revealed that the VIF's for these variables were good. We then performed a final analysis to check the optimal probability of the variables being generated.

I. After creating the ROC curve for the various features, we checked the sensitivity, specificity, and accuracy of the model. We also performed a precision and recall test to see if the model can correctly predict 80% of the cases.

J. We then performed a conversion probability analysis using the sensitivity and specificity metrics. The results of the analysis revealed that the model can correctly predict a conversion rate of 77.59%.

K. The lead score that was obtained from the test set revealed that the model can predict a lead's conversion rate at 83%. This is in line with the expectations of the CEO, who has estimated that the lead conversion rate should be around 80%. The model's sensitivity to the data will help identify the most promising leads.

L. The top 3 variables that contribute for lead getting converted in the model are :

1. Total time spent on website
2. Lead Add Form from Lead Origin
3. Had a Phone Conversation from Last Notable Activity