Lead Scoring Case Study Summary

Problem Statement:

X Education sells online courses to industry professionals. X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.

The company needs a model wherein a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%

Solution Summary:-

Step 1: Reading and Understanding Data.

Read and analyze the data.

Step 2: Data Cleaning:

We dropped the variables that had a high percentage of NULL values in them. This step also included imputing the missing values as and where required with median values in case of numerical variables and creation of new classification variables in case of categorical variables. The outliers were identified and removed.

Step 3: Data Analysis:-

Then we started with the Exploratory Data Analysis of the data set to get a feel of how the data is oriented. We visualized the columns and have dropped some columns due to data imbalance. We also checked the outliers.

Step 4: Creating Dummy Variables:-

Dummy variables are then created.

Step 5: **Train Test Split**-

The next step was to divide the data set into test and train sections with a proportion of 70-30% values.

Step 6: Feature Scaling

We used the Standard Scaler to scale the original numerical variables.

Step 7: Feature Selection using RFE

Using RFE, we selected the top 15 variables.

Step 8: Building Initial Model

Building the initial model by dropping columns with VIF more than 2.0 and also dropping columns with p values more than 0.05

Step 9: Checking Model Accuracy and Metrics

We calculated Sensitivity, Recall, Precision and Specificity. We plotted the ROC Curve and also found the optimal cut off.

Step 10: Making Predictions on test set

We used the model to predict on the test set and checked for the final metrics.