

Lead-Scoring Casestudy

Summary

This case study is about an education company named X Education. Who sells online courses to industry professionals. This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The data provided to us gave a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

The Steps used for the analysis are:

1.Data Cleaning: The columns where the missing values were more we deleted those columns. Imputed the columns where the missing values are less. we replaced the option select with a null value since it did not give us much information. we changed few of the null values to 'not provided' so as to not lose much data, But later removed those while making dummies.

2.EDA: EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seem good.

3.Dummy variable Creation: There were many sub-categories for the categorical features, so we created dummies for those and later we deleted the dummies with 'not provided'.

4.Train-Test split: The split was done at 70% and 30% for train and test data respectively.

5. Model Building: We use RFE first with 22 variables. Later we deleted some variables depending on the P-value and VIF (p-value<0.05 and VIF<5 kept).

6.Model Evaluation: A confusion matrix was made. Later on, the optimum cut off value (using ROC curve and) was used to find the accuracy, sensitivity and specificity which came out to be 80.54%, 83.56% and 78.69% respectively.

7. Prediction: Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 79.97%, 86.83% and 75.83% respectively.

8.Precision – Recall: This method was also used to recheck and a cut off of 0.4 was found with Precision around 68% and recall around 87% on the test data frame.

It was found that the variables that mattered the most in the potential buyers are:

- What is your current occupation--Working Professional
- What is your current occupation--Businessman
- What is your current occupation--Student
- Lead Source--Welingak Website.
- Lead Origin--Lead Add Form.
- Total Time Spent on Website.

Top features to focus to increase Conversion Rate:

- Lead Source--Direct Traffic.
- Lead Source--Google.
- Lead Source--Organic Search.
- Lead Source--Referral Sites.
- Last Activity--Email Bounced.