

LEAD SCORING CASE STUDY

Summary Report

X Education Company sells online courses to industry professionals. Despite getting a lot of leads, it has poor conversion rate, they acquire 100 leads in a day, only about 30 are converted.

The business goal is identifying the most promising leads that are converted into paying customers. The company requires building a model by assigning lead score to each of the leads such that customers with higher lead score have a higher conversion chance and vice-versa.

Following steps were followed throughout the process:

1. Read and Inspect data: Finding out the shape, info, data types, duplicates present in the dataset.
2. Data Cleaning :
 - Handle the "Select" level that is present in many of the categorical variables.
 - Drop columns that are having high percentage of missing values. Check all the columns before dropping them.
 - Check the number of unique categories in each categorical column. Here you may need to do something.
 - For the columns with less percentage of missing, use some imputation technique.
 - Finally check the percentage of rows retained in data cleaning process.
3. Data Preparation:
 - First step is by doing Exploratory Data Analysis (Univariate, Bivariate) and subsequently by drawing a heatmap to calculate correlation between variables.
 - Create dummies for all categorical columns
 - Perform train-test split.
 - Perform scaling
4. Model Building:
 - By using a mixed approach and using RFE we will manually select or attain 15 variables based on the p-value and VIF of the variables. The variables with VIF less than 5 and p-value less than 0.05 were kept.
5. Evaluating the Model :
 - We will start out by creating a confusion matrix. It is then followed by finding the optimum cut off value by using ROC Curve to find the accuracy, sensitivity and specificity.
6. Prediction on test set:
 - Prediction on test data frame and with an optimum cut-off just under 0.4
7. Precision-Recall:
 - The value of precision and Recall on test data set is 73% and 76.5% respectively.
8. Conclusion:
 - The variables that affect most in conversion of a visitor into a paying customer are
 - Total time spent on website
 - Lead Origin
 - Lead Source