

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 you 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.

Solution Summary:

Step1: Reading and Understanding Data: Read and inspected the data.

Step2: Data Cleaning:

- First step to clean the dataset we chose was to drop the variables having unique values (Prospect ID etc.).
- Then, there were few columns with value 'Select' which means the leads did not choose any given option. We changed those values to Null values.
- We dropped the columns having NULL values greater than 40%.
- Next, we removed the imbalanced and redundant variables. 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.
- Also, in one column was having identical label in different cases (like Google and google). Those data were corrected.

Step3: Dummy Variables Creation:

- We created dummy variables for the categorical variables.
- Removed all the repeated and redundant variables

Step5: Test Train 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 Rescaling:

- We have used the Standard Scaling to scale the original numerical variables.

Step 7: Model Building:

- Using the Recursive Feature Elimination, we went ahead and selected the 30 top important features.
- We built model using selected features and all features and evaluated them.

Step 8: Conclusion:

The lead score calculated in the test set of data shows the conversion rate of 88% on the final predicted model which clearly meets the expectation of CEO has given a ballpark of the target lead conversion rate to be around 80%.