

Summary

Problem Statement:

Help X Education to select the most promising leads (the leads that are most likely to convert into paying customers) so that target lead conversion rate could be around 80%.

Utilizing past 9000 data points build a model and assign a lead score (between 0 to 100) 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 Approach:

1. EDA:

- Check was done on % of null value and we dropped columns with more than 45% missing values.
- We worked on numerical variable, outliers and dummy variables.
- Bucketing the values which have less percentage of value_count as compared to other values in the parameter.

2. Train-Test split & Scaling :

- The split was done at 70% and 30% for train and test data respectively.
- We used standard scaler approach for scaling.

3. Model Building

- RFE was used for feature selection with 15 relevant variables.
- Rest of the variables were removed manually depending on the VIF values and p-value.
- A confusion matrix was created, and overall accuracy was checked which came out to be 80.80%.
- We assigned the lead score which is **probability*100**, this will help to identify hot leads.

4. Model Evaluation

- **Sensitivity – Precision**

- Prediction on **Training Data**

- The optimum cut off value was found using ROC curve. The area under ROC curve was 0.89.
 - After Plotting we found that optimum cutoff was **0.35** which gave
 - Precision: 72.09%
 - Sensitivity: 80.02%
 - Accuracy: 80.80%
 - Specificity: 81.12%

- Prediction on **Test Data**

- Precision: 72.82%
 - Sensitivity: 80.31%
 - Accuracy: 81.20%
 - Specificity: 81.75%

CONCLUSION:

The order of the parameters (high to low) that will impact the Lead conversion:

- What is your current occupation_Working Professional
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- What is your current occupation_Other
- Last Activity_SMS Sent
- Total Time Spent on Website
- What is your current occupation_Unemployed
- Last Notable Activity_Modified
- Lead Source_Google
- Lead Source_Referral Sites
- Lead Source_Organic Search
- Do Not Email_Yes
- Lead Source_Direct Traffic