Leads Scoring Case Study

By Parthajit Basu

Steps involved while creating the model

1) Data Cleaning
2) Data Transformation
3) Data Preparation
4) Model Building
5) Model Evaluation

Data Cleaning

- First we tried to check Null values in columns
- Then we tried to check which all columns have “Select” key word and converted them to NULL
- Later tried to get percentage of Null in all columns
- Later we tried to replace null values in columns based on highest mode value in each column

- We tried few approaches to trim down redundant values in columns to certain category

Data Transformation

- Checked the outliers and created bins for them.
- Removed all the redundant and repeated columns.
- Changed the multi category labels into dummy variables and binary variables into ‘0’ and ‘1’.

Data Preparation

- Split the dataset into train and test dataset
- Scaled necessary date set

Model Building

- We created our model with 15 variables which is standard process
- We tried to check confusion matrix, accuracy, VIF. Based on VIF score deleted columns which have more than 5 VIF points.
- Rebuilding the model again and tried to find confusion matrix, Accuracy, VIF, Sensitivity, Specificity
- Tried to check ROC curve and tried to find cutoff value

Model Evaluation

- We did model evaluation on the test set like checking the accuracy, recall/sensitivity to find how the model is performing

- We found the score of accuracy and sensitivity from our final test model is totally in acceptable range.

- We have given lead score to the test dataset for indication that high lead score are hot leads and low lead score are not hot leads.