Lead Score Case Study – Summary

The Lead Score Case Study aimed at using the logistic regression model to identify factors influencing the lead conversion rate of perspective customers for X Education.

The target of the assignment was to give a Lead Score between 1 – 100 to each lead and increase the conversion rate.

The problem was approached in following steps:

1. Inspecting the Data Frame
2. Exploratory Data Analysis
3. Data Preparation
4. Building The Model
5. Feature Scaling
6. Building the Correlation Metrix
7. Build the Model
8. Plotting the ROC Curve
9. Finding the Optimal Cutoff Point
10. Making Predictions on Test Set

The Exploratory Data Analysis steps several unwanted columns were removed and data structure was corrected by imputing the Null values. Once done, the data was split into Train-Test set with 70% random values in Train Set.

After converting the variables to binary and assigning the dummy values to variables where required, the data was scaled and RFE method was used to build the initial model. Once the variables were identified, further cleaning was done using the VIF method. The overall accuracy of the model using these steps was found to by 78%.

Further, the ROC curve was plotted and optimal probability value was found to be 0.25. For further verification, the confusion matrix was designed. The model was finally tested on the Test set and an overall accuracy of 78% was found.